PCT/GB 9 6 / 0 1 3 4 1

The Patent Office

0 5 June 1996

The Patent Office
Cardiff Road
Newport
Gwent
NP9 1RH



I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation and Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated 19th June 1996

For official use


ORIGINS 512331A-9 [illegible]
PO1/7700 25.00

Your reference PP/KP/1172

- 6 JUN 1995

**9511439.3**

***Notes***

*Please type, or write in dark ink using CAPITAL letters. A prescribed fee is payable for a request for grant of a patent. For details, please contact the Patent Office (telephone 071-438 4700).*

*Rule 16 of the Patents Rules 1990 is the main rule governing the completion and filing of this form.*

*❷ Do not give trading styles, for example, 'Trading as XYZ company', nationality or former names, for example, 'formerly (known as) ABC Ltd' as these are not required.*

***Warning***

*After an application for a Patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977 and will inform the applicant if such prohibition or restriction is necessary. Applicants resident in the United Kingdom are also reminded that under Section 23, applications may not be filed abroad without written permission unless an application has been filed not less than 6 weeks previously in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction revoked.*

The Patent Office

# Request for grant of a Patent

## Form 1/77

**Patents Act 1977**

## ❶ Title of invention

1 Please give the title of the invention: GENE PRODUCT AND METHOD

## ❷ Applicant's details

❑ **First or only applicant**

2a If you are applying as a corporate body please give:

Corporate name: ISIS INNOVATION LIMITED

Country *(and State of incorporation, if appropriate)*: United Kingdom

2b If you are applying as an individual or one of a partnership please give in full:

Surname

Forenames

2c **In all cases**, please give the following details:

Address: 2 SOUTH PARKS ROAD
OXFORD

UK postcode *(if applicable)*: OX1 3UB

Country: England

ADP number *(if known)*: 03998564001

**2d, 2e and 2f:** *If there are further applicants please provide details on a separate sheet of paper.*

❑ **Second applicant** ***(if any)***

2d If you are applying as a corporate body please give:

Corporate name

Country *(and State of incorporation, if appropriate)*

2e If you are applying as an individual or one of a partnership please give in full:

Surname

Forenames

2f **In all cases**, please give the following details:

Address

UK postcode *(if applicable)*

Country

ADP number *(if known)*

❸ *An address for service in the United Kingdom must be supplied*

*Please mark correct box*

## ❸ Address for service details

3a Have you appointed an agent to deal with your application?

Yes [X] No [ ] ➡ ***go to 3b***

⬇ ***please give details below***

Agent's name: Stevens, Hewlett & Perkins

Agent's address: 1 Serjeants' Inn
Fleet Street
London

Postcode: EC4Y 1LL

Agent's ADP number: 1545003

**3b:** *If you have appointed an agent, all correspondence concerning your application will be sent to the agent's United Kingdom address*

3b If you have not appointed an agent please give a name and address in the United Kingdom to which all correspondence will be sent:

Name

Address

Daytime telephone number *(if available)*

Postcode

ADP number *(if known)*

## ❹ Reference number

4 Agent's or applicant's reference number *(if applicable)* PP/KP/1172

## Claiming an earlier application date

5 Are you claiming that this application be treated as having been filed on the date of filing of an earlier application?

*Please mark correct box*

Yes ☐ No ☒ ➡ ***go to 6***

↓ ***please give details below***

☐ number of earlier application or patent number

☐ filing date *(day month year)*

☐ and the Section of the Patents Act 1977 under which you are claiming

*Please mark correct box*

15(4) (Divisional) ☐ 8(3) ☐ 12(6) ☐ 37(4) ☐

## ❻ Declaration of priority

*❻ If you are declaring priority from a PCT Application please enter 'PCT' as the country and enter the country code (for example, GB) as part of the application number.*

*Please give the date in all number format, for example, 31/05/90 for 31 May 1990.*

6 If you are declaring priority from previous application(s), please give:

| Country of filing | Priority application number *(if known)* | Filing date *(day, month, year)* |
|---|---|---|
| | | |

❼ *The answer must be 'No' if: any applicant is not an inventor there is an inventor who is not an applicant, **or** any applicant is a corporate body.*

## Inventorship

7 Are you (the applicant or applicants) the sole inventor or the joint inventor?

*Please mark correct box*

Yes [ ] No [X] ➡ ***A Statement of Inventorship on Patents Form 7/77 will need to be filed (see Rule 15).***

❽ *Please supply duplicates of claim(s), abstract, description and drawing(s).*

## ❽ Checklist

8a Please fill in the number of sheets for each of the following types of document contained in this application.

| | | | |
|---|---|---|---|
| Continuation sheets for this Patents Form 1/77 | | | |
| Claim(s) | 3 | Description | 21 |
| Abstract | | Drawing(s) | 15 |

8b Which of the following documents also accompanies the application?

*Please mark correct box(es)*

| | |
|---|---|
| Priority documents *(please state how many)* | |
| Translation(s) of Priority documents *(please state how many)* | |
| Patents Form 7/77 – Statement of Inventorship and Right to Grant *(please state how many)* | |
| Patents Form 9/77 – Preliminary Examination/Search | |
| Patents Form 10/77 – Request for Substantive Examination | |

❾ *You or your appointed agent (see Rule 90 of the Patents Rules 1990) must sign this request.*

## ❾ Request

I/We request the grant of a patent on the basis of this application.

***Please sign here*** ➡

Signed Stevens, Hewlett + Perkins Date 06. 06. 95 (day / month / year)

STEVENS, HEWLETT & PERKINS

*A completed fee sheet should preferably accompany the fee*

***Please return the completed form, attachments and duplicates where requested, together with the prescribed fee to either:***

❑ ***The Comptroller***
***The Patent Office***
***Cardiff Road***
***Newport***
***Gwent***
***NP9 1RH***

***or***

❑ ***The Comptroller***
***The Patent Office***
***25 Southampton Buildings***
***London***
***WC2A 1AY***

## GENE PRODUCT AND METHOD

### Introduction

The present invention relates to proteins, polypeptides, nucleic acid fragments, antibodies and related products and to their use in medicine and agriculture, for instance in diagnosis and therapy. More particularly the invention relates to a gene or genes which control the sex of the embryos of birds and to their use in ascertaining the sex of cells, embryos and tissues and controlling the sex of the progeny of birds

Much of our understanding of sex determination comes from three, extensively studied, model systems. In two of these, the fruitfly *Drosophila melanogaster* and the nematode *Caenorhabitis elegans*, it is the ratio of X chromosomes to autosomes that initiates sexual differentiation [Hodgkin, 1992 #435]. In the mouse a single gene, *sry*, located on the Y chromosome provides the impetus for male development; a pattern that is thought to be conserved throughout the mammals [Koopman, 1991 #319 Foster, 1992 #323].

At the genetical level these three species employ very different molecular mechanisms, not only to control sex determination itself but to accommodate the differing dosages of genes that result from the males possessing a single X and the female two X chromosomes. These basic differences are largely due to the independent evolution of the three mechanisms and strongly suggests that other means of sex determination will have evolved elsewhere in the animal kingdom.

One class in which little is known about sex determination is the birds. They exhibit female heterogamety which means that the female has Z and W sex chromosomes and the male ZZ. This immediately suggests that sex determination in this class has an independent origin to that of their sister class, the mammals where it is the male that is heterogametic. Furthermore, it has been shown that whilst female mammals inactivate one of their X chromosomes as a method of dosage compensation [Grant, 1988 #449], this does not seem to be a device employed by birds [Baverstock, 1982 #102].

However, similarities do exist between between the birds and mammals. The W chromosome, like the Y chromosome is usually smaller than its partner, and is als characteristically heterochromatic in appearance [Christidis, 1990 #440]. The main exceptions to this rule are found in the 'primitive' representatives of both classes: the monotremes and the ratites where the morphological differences between the sex chromosomes are poorly defined [Graves, 1987 #408; Tagaki, 1972 #205].

The heterochromatization of the W and Y results from the replacement of functional genetic loci with junk DNA sequences. This process is thought to be a consequence of a suppression of recombination that has arisen to ensure that genes vital to the development of the heterogametic sex remain linked on the Y or W chromosome [Charlesworth, 1991 #407]. As a result only a few genes such as *Ubely* [Kay, 1991 #328; Mitchell, 1991 #327], *Zfy* [Page, 1987 #322] and *SRY* itself remain on the mammalian Y chromosome. A similar situation is thought to prevail on the avian W chromosome where the presence of any functional genes has yet to be demonstrated, although it does possess vast arrays of repetitive elements [Tone, 1982 #178; Griffiths, 1990 #77] .

A further similarity in sex determination in birds and mammals is that the development of the male phenotype appears crucially dependent on the appearance of the testis. The female phenotype is the result of the 'default pathway'. For mammals this was first demonstrated by Jost (1947) who grafted an embroyonic testis into genetically female rabbit embryos prior to sex determination. This was sufficient to allow the development of functional males. The same experiment has been carried out on chick embryos [Stoll, 1978 #451] with comparable results.

Once the testis has formed, the process of masculinization is adopted by the testicular hormones. The genetical switch that initiates testis determination is known to be *SRY* in mammals [Koopman, 1991 #319]. In birds, there appears to be no *SRY* homologue on the W chromosome [Griffiths, 1991 #78], although this is unsurprising given the separate evolution of sex determination in the two classes.

The only other pertinent evidence on the genetics of avian sex determination come from reports of chickens with abnormal chromosome complements. Table 1 shows data from [Crew, 1954 #450] and [McCarrey, 1979 #448] on the phenotypes of the aneuploids so far described. These results suggest that the presence of the W chromosome in the aneuploid AA ZZW and the polyploid AAA ZZW has not acted as a dominant determinant of the female phenotype. This may mean that sex in birds may be determined more by the autosome to Z ratio, as in *Drosphila* and *C. elegans*. However, a ZO aneuploid which could confirm this hypothesis has yet to be described.

It must also be born in mind that XXY kangaroos, where *SRY* is thought to be the key male determining switch, exhibit both male and female characteristics [Graves, 1987 #408]. This suggests that the limited aneuploid data that is available for birds should be interpreted with some caution.

To conclude, the genetic mechanism that controls sex determination in birds has not yet been elucidated. In this paper we suggest that a gene we have termed *CHD-W* (*C*hromodomain-*H*elicase-*D*NA binding on the*W* chromosome) alone or acting in conjunction with a closely related gene *CHD-1A* (*C*hromodomain-*H*elicase-*D*NA binding 1 *A*vian) initiates female development in birds.

It is believed that the all birds such as chickens and other species of commercial significance, will have two or more genes of the CHD type which will have a nucleotide sequence similar to the nucleotide sequences shown in fig. 5, fig. 7 and fig. 8 and that the gene products will be proteins which are crucuial to the determination of the sex of the organism. One of these genes will be located on the W chromosome and the other on an autosome or Z chromosome.

It will be understood that the exact sequence of the two genes will vary between species and between individuals of the same species at least at the nucleotide level and often also at the protein level. Partial sequences of the chicken genes are shown in fig. 5, fig. 7 and fig. 8. The gene or protein which contains sequence corresponding to those in fig. 5, fig. 7 and fig. 8 will hereafter be referred to as an CHD-gene and proteins and fragments thereof, polypeptides, nucleic acids and fragments thereof and oligonucleotides containing part of a CHD gene will hereafter be referred to as CHD-proteins, CHD-nucleic acids and so on.

The present invention therefore provides a CHD-protein or a fragment thereof or polypeptide comprising a CHD-gene or a part thereof, subject to the proviso below.

The present invention also provides a protein or a fragment thereof or a polypeptide containing a mimetope of an epitope of a CHD-protein or fragment thereof of polypeptide containing a CHD-gene or a part thereof, subject to the proviso below. Such proteins, fragments and polypeptides are hereafter referred to as CHD-mimetope proteins or fragments thereof and mt-mimetope polypeptides.

The present invention also provides a CHD-nucleic acid or a fragment therof or oligonucleotide comprising a CHD-gene, or a part thereof subject to the proviso below.

In a particular aspect the present invention provides a single or double stranded nucleic acid comprising the CHD-gene of a bird or a part thereof of at least 17 contiguous nucleotide bases or base pairs, or a single or double stranded nucleic acid hybridizable with the CHD-gene of a bird, or part thereof of at least 17 contiguous nucleotide bases or base pairs, subject to the proviso below.

The invention further provides a nucleic acid or fragment thereof or an oligonucleotide encoding a CHD-protein or fragment thereof or a polypeptide comprising a CHD-gene or a

part thereof or a CHD-mimetope protein or a fragment thereof or CHD-mimetope polypeptide subject to the following proviso. These nucleic acids, fragments and oligonucleotides may hav sequences differing from the sequences of CHD-nucleic acids, fragments and oligonucleotides due to alternative codon useage and/or encoding alternative amino acids sequences or mimetopes.

The present invention does not, however extend to any known protein or fragment thereof or polypeptide or nucleic acid or fragment thereof or oligonucleotide containing a CHD-gene related sequence such as the *Saccharomyces cerivisiae SNF2/SWI2* gene, *Drosophila polycomb* and HP1 genes described below, insofar as that protein or fragment, polypeptide, nucleic acid or fragment or oligonucleotide is known *per se*.

The amino acid sequence of the CHD-gene has similarities to the chromobox and helicase motifs of a number of known genes known to be involved in the remodelling of chromatin. This suggests that the CHD-protein of the present invention may have a regulatory function involving chromatin remodelling. However, none of these genes contain the chromobox and the helicase of the CHD-gene which are conserved in conjunction, at least in the chicken, great tit, mouse and yeast but are not conserved in conjunction in the the sequences of chromatin remodelling proteins not associated with sex determination at least at the stage of testis formation in birds. A protein having chromatin remodelling capacity but lacking these characteristic motifs is therefore outside the scope of the present invention.

In addition there are certain residues in the amino acid sequence of the chromobox and those residues immediately downstream thereof, of the CHD-gene which are also conserved at least between those found in the chicken, great tit, mouse and yeast but are not conserved in the the sequences of chromatin remodelling proteins not associated with sex determination at least at the stage of testis formation in birds. Any one of these conserved residues is therefore considered characteristic of the CHD-gene proteins of the present invention. A protein having chromatin remodelling capacity and a helicase motif but lacking all or most of these characteristic amino acid residues in the chromobox motif is therefore outside the scope of the present invention.

The characteristic amino acids residues are shown in the alignment in fig. 11, which is described in more detail below. When aligned with the illustrated sequences as shown, these residues fall at positions, 11,12, 20, 27, 34 inside the chromobox and 3, 6, 8, 12-15, 16 immediately downstream.

The nucleotide base sequence of the CHD-gene includes bases which encode the chromobox and helicase motifs of chromatin remodelling proteins as described above. However, the base sequence of the CHD-nucleic acids of the gene will include codons specifying both chromobox and helicase motifs and the former will have codons specifying one or more of the characteristic amino acid residues described above and/or will be hybridizable with a sequence that controls the sex determination of birds under conditions which substantially prevent hybridization to other sequences in birds that do not have these characteristics.

Preferably the CHD-nucleic acids of the invention encode a chromobox and a helicase and one or more, preferably all, of the characteristic chromobox amino acid residues and meet the above hybridization requirements.

Fragments of CHD-nucleic acids according to the present invention will likewise contain codons specifying the chromobox and helicase motifs or including at least part of either of these motifs or CHD-gene adjacent to the codons encoding these features and/or will be hybridizable with a sequence that controls the sex determination of birds under conditions which substantially prevent hybridization to other sequences in birds that do not have these characteristics.

Oligonucleotides containing the CHD-gene or a part thereof according to the present invention may contain codons specifying the chromobox or helicase motifs or including at least part of these motifs or CHD-gene but this is not essential. However all such oligonucleotides of the invention must be capable of hybridizing with a sequence or sequences that control the sex determination of birds, preferably under conditions which substantially prevent hybridization with any sequence not associated with sex determining sequence.

The sex determining sequences referred to herein is the sequence which contains the CHD-gene and which encodes a factor which when expressed at the appropriate stage and level during embryo development may result in testis formation and susequent growth of the embryo as a male. It may alternatively refer to the sequence which encodes a factor which when expressed at the appropriate stage and level during embryo development prevents testis formation and results in the subsequent growth of the embryo as a female.

The hybridization conditions referred to above which prevent unwanted hybridization with sequences not associated with the sex determining gene will depend to some extent on the length of the nucleic acid, fragment or oligonucleotide of the invention tested. Thus for instance lower stringency will be sufficient to secure hybridization to sequences associated with the sex

determining gene whilst preventing unwanted hybridization when the nucleic acid or fragment is several thousand nucleotide base pairs in length than for a fragment of only a few hundereds bases or an oligonucleotide of from 17 bases up to a few tens or hundereds of bases. With the smallest oligonucleotides and fragments of the invention hybridization conditions will be such that only complete complementarity between the oligonucleotide and or fragment and the sequences associated with the sex determining gene will result in hybridization.

Preferred nucleic acids and fragments of the invention will only hybridize selectively to the sequences associated with the sex determining gene or genes under conditions requiring at least 80%, for instance 85, 90 or even 95% mor preferably 99% complementarity. Yet more preferred nucleic acids and fragments of the invention are those having a sequence corresponding exactly to that of those illustrated in fig. 5, fig. 7 and fig. 8 although the nucleotide sequences by be longer or shorter than those illustrated and or may contain normally intronic sequences associated with these sequences

The invention particularly provides an oligonucleotide, polypeptide, nucleic acid or protein comprising the entire sequence of the CHD-gene of a bird and more preferably comprising the entire amino acid or nucleotide sequence of the chicken as set out in any one of figs 1, 3, 5, 7, 8, 9, 10, 11.

The nucleic acids hybridizable with the CHD-gene of a bird are preferably hybridizable under moderate, or more preferably, high stringency conditions as defined below:

**Moderate stringency:**

| | |
|---|---|
| Buffer: | 2 x SSC |
| Temp: | 50°C |
| Annealing period: | 6-8hrs |

**High stringency:**

| | |
|---|---|
| Buffer: | 1 x SSC |
| Temp: | 65°C |
| annealing period: | 6-8hrs |

Moderate stringency as defined above corresponds with about 75% homology. High stringency as defined above corresponds with about 90% homology. 1 xSSC is 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0.

Preferably the portion of the nucleic acid corresponding to or hybridizable with the CHD-gene is at least 20, more preferably at least 30, 40 or 60 and most preferably 100 or more nucleotide bases in length.

The nucleotide strands of the invention may be single or double stranded DNA or RNA. DNA's of the invention may comprise coding and/or non-coding sequences and/or transcriptional and or translational start and/or stop signals and/or regulatory, signal and/or control sequences such as promotors, enhancers and/or polyadenylation sites, endonuclease restriction sites and/or splice donor and/or acceptor, in addition to the CHD-gene sequence. Included within the DNA's of the invention are genomic DNA's and complementary DNA's (cDNA's) including functional genes or at least an exon containing the CHD-gene. They may also contain non-coding sequences such as one or more introns. Single stranded DNA may be the transcribed strand or the non-transcribed (complementary) strand. The nucleic acids may be present in a vector, for instance a cloning or expression vector, such as a plasmid or cosmid or a viral genomic nucleic acid. RNA's of the invention include unprocessed and processed transcripts of DNA, messenger RNA (mRNA) containing th CHD-gene and anti-sense RNA containing a sequence complementary to the CHD-gene.

Nucleic acids of the present invention are particularly useful as primers for polymerase chain reactions (PCRs) conducted to ascertain the sex of a bird as defined below. They may also be used to express proteins or fragments or polypeptides corresponding to the whole or a part of a CHD-protein (whether or not containing a CHD-gene) or as probes in hybridization experiments. As used herin the term "fragments" used in connection with proteins is intended to refer to both chemically produced and recombinant portions of proteins.

The CHD-proteins and fragments thereof and polypeptides containing the CHD-gene or a part thereof and CHD-mimetope proteins and fragments thereof and CHD-mimeotope polypeptides of the invention are useful in immunodiagnostic testing and for rasing antibodies such as monoclonal antibodies for such uses. Antibodies against such proteins and fragments and polypeptides as well as fragments of such antibodies (which antibody fragments include at least one antigen binding site) including chemically derived and recombinant fragments of such antibodies, and cells, such as eukaryotic cells, for instance hybridomas and prokaryotic recombinant cells capable of expressing and, preferably secreting antibodies or fragments thereof against such proteins or fragments, also form part of the present invention.

The nucleic acids of the invention may be obtained by conventional means such as by the recovery from organisms using PCR technology or hybridization probes, by *de novo* synthesis

or a combination thereof, by cloning the CHD-nucleic acids described below or a fragment thereof or by other techniques well known in the art of recombinant DNA technology.

Proteins and fragments thereof and polypeptides of the invention may be recovered from cells of organisms expressing a CHD-gene or generated by expression of a CHD-gene or coding sequence contained in a nucleic acid of the present invention in an appropriate expression system and host, or obtained by *de novo* synthesis or a combinmation thereof, by techniques well known in the art of recombinant DNA technology. The proteins, fragments thereof and polypeptides of the invention will contain naturally occuring L-a-amino acids and may also contain one or more non-naturally occurring a-amino acids having the D- or L-configuration

Antibodies may be obtained by immunization of a suitable host animal and recovery of the antibodies, by culture of antibody producing cells obtained from suitably immunized host animals or by *in vitro* stimulation of B-cells with a suitable CHD-protein, fragment or polypeptide or CHD-mimetope, protein, fragment or polypeptide and culture of the cells. Such cells may be immortalized as necessary for instance by fusion with myeloma cells. Antibody fragments may be obtained by well known chemical and biotechnological methods.

All these techniques are well known to practitioners of the arts of biotechnology. Reference may particularly be made to the well known text book "Molecular cloning: A laboratory manual" 2nd Edition (Eds Sambrook, J., Fritsch, E.F. and Maniatis, T.), (Cold Spring Harbour Laboratory, New York, 1989), hereafter referred to as "Maniatis".

The invention further provides the use of a nucleic acid, protein, polypeptide, antibody, or antibody producing cell as hereinbefore defined including the SNF2/SWI2, polycomb and HP1 or other chromobox or helicase containing protein for ascertaining the sex of a cell or organism of a bird or for isolating nucleic acids useful in ascertaining the sex of a bird and for instituting single sex breeding programmes.

Knowledge of the chicken or great tit sex determining gene or genes can be used to isolate the equivalent gene or genes from other birds. Once isolated from a particular species, this gene or genes and its sequence can be used in two types of application:

1. The construction of sequence based sexing tests which can be applied to embryos, tissues ant other biological materials containing nucleic acids.

2. The genetic modification of the germ line of birds to create breeding systems that produce offspring statistically biased towards one sex or of one sex only (single sex breeding systems).

A particularly preferred technique for ascertaining the sex of a bird in accordance with the invention involves the use of an oligonucleotides as primers in a PCR, for instance as follows:

A cell or cells or remains thereof are obtained, for instance by surgical removal from an embryo or from the quill of a feather, and the DNA is released by a crude lysis prcedure for instance using a detergent or by heating. Primer olignucleotides of the invention are used to initiate a conventional PCR in order to amplify W chromosome linked CHD-related DNA from the cells. The products of the PCR are analysed by agarose gel electrophoresis and detected using labelled probes or by visual inspection. The presence of amplified DNA indicates the presence of a *CHD-W* gene in the cells and thus, in birds, that the cell(s) were female.

This technique may be applied for instance to identify the sex of embryos or adults for subsequent breeding programs, or to control the sex of the progeny of breeding stock for commercial exploitation (by selection of the breeding stock or by slaughter or termination of animals of undesired sex).

The oligonucleotide primers for ascertaining or controlling sex in one species may also be used to ascertain or control sex in another species since hybridization of the primers to the CHD-gene of the other species will still serve to amplify the species-specific sequences.

Techniques for conducting such determinations are well known in the art of recombinant DNA technology.

In one aspect the present invention provides a process for isolating a W-chromosome specific sequence associated with the *CHD-W* gene of a a bird which comprises probing a genomic library from a female of the species preferably of W chromosome sequences, for instance of lambda phage, cosmid or YAC library or cDNA library constructed from a tissue expressing the gene, with a probe comprising a nucleic acid, fragment or oligonucleotide of the invention as hereinbefore defined and a detectable label under high or moderate stringency.

Using the newly isolated subclone, Southern blots are performed on male and female DNA of the species of interest at high stringency to confirm that the correct clone has been isolated. The CHD-gene probe should give a female specific signal (other male/female shared bands may also be present at lesser intensities). The subclone is sequenced using standard methods and primers suitable for PCR chosen from the sequence so identified.

Alternatively, other approaches to cloning the sequences related to the sex determining ge could be used such as PCR methods using "degenerate" oligonucleotides. (For methods in PCR see, for example, "PCR Protocols - a Guide to Methods and Application"; edited by M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White; published by Academic Press, Inc.).

Preferably the probe is *CHD-1A* or *CHD-W* or a fragment thereof or a nucleic acid or fragment or oligonucleotide having a sequence exactly as set out in Fig. 5, fig. 7 or fig. 8 for the chicken. Techniques for forming a genomic or cDNA library and for probing and detecting the detectable label and isolating the nucleic acid identified by the probe are well known in the art of biotechnology and recombinant DNA manipulation. The process may be conducted for instance using a probe having the chicken sequence such as the *CHD-W* sequence to identify and isolate the corresponding sequence from another bird such as Turkey. The thus-identified sequence can then be used to generate primers for PCR which in turn can be used to ascertain the sex of an individual or of cells, tissues, embryos or ovaries of the bird. This will permit experiments to ascertain sex to be conducted and controlled sex breeding of the bird as described below.

The isolated nucleic acid, fragment or oligonucleotide may thereafter be amplified, cloned or sub-cloned as necessary. The invention further provides a process for detecting the sex of an individual bird or of cells, tissues, embryos, foetuses or ovaries or a bird, comprising conducting a polymerase chain reaction using DNA from the individual, cell, tissue, embryo or ovary as template and a nucleic acid, fragment or oligonucleotide of the invention as primer. Preferably the nucleic acid, fragment or oligonucleotide of the invention used as primer is *CHD-W* or *CHD-1A* or a part thereof and has a sequence corresponding exactly to the chicken sequence in fig. 5, fig. 7 or fig. 8 or a part thereof or is a nucleic acid, fragment or oligonucleotide which is a W-chromosome specific sequence associated with the sex determining gene or genes of a bird of the same species as the individual cell, tissue, embryo, foetus or ovary whose sex is to be ascertained. The W-chromosome specific sequence associated with the sex determining gene or genes of the bird involved may itself have been obtained by the process of isolation and amplification or cloning described above.

The identification of the sex determining gene or genes according to the present invention raises the possibility of controlling the sex of progeny of commercially important animals such as chickens, turkeys and other avians. This will be valuable in many aspects of animal breeding and husbandry such as where one sex has more desirable characteristics, for instance only female progeny are desired for egg-laying breeds of chicken. The economic advantages of single sex breeding programmes and strategies for instituting these are described for instance in

"Exploiting New Technologies in Animal Breeding; Genetic Developments", (Eds. Smith, C., King, J.Q.B. and McKay, J.C.), (Oxford University Press, Oxford, 1986).

The nucleic acids making up all or part of the sex determining gene, from the same or different animal species, can be introduced into any early embryo through established transgenic technology. This latter includes microinjection of DNA into pronuclei or nuclei of early embryos, the use of retroviral vectors with either early embryos or embryonic stem cells, or any transformation technique, (including microinjection, electroporation or carrier techniques) into embryonic stem cells or other cells able to give rise to functional germ cells. These procedures will allow the derivation of individual transgenic animals (founder transgenics) or chimeric animals composed in part of cells carrying the introduced DNA. Where the functional germ cells of the founder transgenic or chimeric animal carry the introduced DNA it will be possible to obtain transmission of the introduced DNA to offspring and to generate lines or strains of animals carrying these DNA sequences.

The nucleic acids making up part or all of the coding sequence of the sex determining gene, or derivatives of it, may be introduced in combination with its own regulatory sequences (promoter/enhancers etc.) or regulatory sequences from another gene, the whole making the "construct", to give expression from the construct at an appropriate developmental stage and tissue location critical to sex determination in the bird species under consideration. For example, in the chicken this would be between 6 and 7 days post lay.

## Materials and Methods

### Isolation of pGT-W, pGT1.7 and pGT8

A great tit (*Parus major*) library was constructed from genomic DNA, partially restricted with *Mbo*I, and the lFixII vector (Stratagene). The library was screened at high stringency with the 724bp probe (GT-W) cloned from a W chromosome specific polymerase chain reaction (PCR) product derived from the great tit [Griffiths, 1993 #379]. Positive plaques were subject to two rounds of purification. Clone lGT2 contained an insert of 9.6kb that hybridized strongly to the probe sequence. The insert was subcloned as two *EcoRI* fragments of 1.7kb (pGT1.7) and 8kb (pGT8) into *EcoRI* cut pT7/T3 (Pharmacia).

### Isolation of CHD genes from the chicken

Two chicken cDNA libraries were screened. The first was a mixed sex chick stage 10-12 cDNA library in lZapII which had been reamplified on 2 occassions This library was provided by Dr I.J. Mason. The second library was constructed from mixed sex, 10 day chick mRNA. Total RNA was extracted using a guanidine thiocyanate based technique [Koopman, 1993

#550] and mRNA isolated using a Promega PolyATtract system 1000. A lZapII library was constructed using a Stratagene ZAP-cDNA synthesis kit.

Plaques ($2x10^5$) from the stage 10-12 day library were screened at moderate stringency with a subcloned 433bp HindIII/SacI fragment from pGT8 that contained the 123bp region with identity to the mouse *CHD-1* gene [Delmas, 1993 #415]. A similar number of plaques from both libraries were screened with bases 122-4132 of *CHD-1A* (see fig. 5 ). The 10 day library was also screened with bases 3763-5007 of *CHD-1A* (see fig. 5). Positive plaques were purified prior to the excission of pBluescript plasmids and cloned inserts insert from lZapII using techniques recommended by Stratagene.

**Sequencing**

All sequencing was carried out using the T7 DNA polymerase/7-deaza-dGTP chain termination sequencing kit from USB. All sequencing unless otherwise specified was carried out in both directions either by subcloning or through exonuclease III deletion with the Promega Erase-a-Base system.

**Southern Blot Analysis and Hybridization**

Genomic DNA was extracted from blood [Griffiths, 1990 #77], digested with the appropriate restriction enzyme and Southern blotted onto Zeta-Probe GT under neutral conditions as described by the manufacturer (Bio-Rad). Prehybridizations and hybridizations were carried out in 0.25M $Na_2HPO_3$/5% SDS at either 65°C (high stringency) or 62°C (moderate stringency). Subsequent washes were carried out for a total of 1 hour in three changes of either 0.5 x SSC (75mM NaCl/7.5mM sodium citrate (pH7.5))/0.1%SDS at 65°C (high stringency) or 1 x SSC/0.1%SDS at 45°C (low stringency).

## Results

The plasmid pGT-W contains a 724bp insert that hybridizes to a 4.9kb fragment only in the female great tit. Its DNA sequence was determined (fig.1) and contains a 457bp open reading frame. A search of the EMBL DNA and protein sequence database found no significant matches. The sequence does contain a simple sequence consisting of a 22bp run of thymidines.

The pGT-W insert was used to probe Southern blots, at low stringency, of *PvuII* restricted genomic DNA of male and female great tit, starling (*Sturnus vulgaris*), jackdaw (*Corvus monedula*), pied wagtail (*Motacilla alba*) and a species of new world flycatcher. These are species that cover the extremes of the passeriforme order according to the recent phylogeny of [Sibley, 1988 #409]. In all but the jackdaw convincing hybridization to a single female specific fragment could be observed. In all, species hybridization to one or more non-sex specific

fragments was also shown. A similar experiment was carried out with a non-passerine, the bee-eater (*Merops apiaster*), and this too resulted in faint hybridization to a female specific fragment and two, somewhat stronger bands, in both sexes.

In order to further investigate the nature of the pGT-W insert we attempted to clone a larger frament of genomic DNA which incorporated this motif. From around 1.5 x$10^5$ plaques from a great tit genomic library, two positives were obtained. After purification one of these gave superior hybridization and was investigated further. The 9.7kb insert was subcloned as pGT1.7 and pGT8 containing 1.7kb and 8 kb respectively. The pGT1.7 was sequenced in its entirety and approximately 2.8kb of the sequence of pGT8 was determined both in only a single direction. A 723bp region, starting 133bp from the 5' end of pGT8 had a sequence that corresponded exactly to the pGT-W insert (fig. 2).

The sequences derived from these subclones were used to search the EMBL database using the FASTA algorithms (GCG, Wisconsin package vers 7.3). A region of 123bp, starting 994bp from the 5' end of pGT8, showed a 79% nucleotide sequence identity to bases 3855-3977 of the mouse *CHD-1* gene (fig. 3)[Delmas, 1993 #415]. This corresponds to an 88% identity at the amino acid level.

Southern blots of *PvuII* digests of genomic DNA from male and female chicken and lesser black-backed gull (*Larus fuscus*) were probed at low stringency with a 433bp SacI/HindIII fragment of pGT8 that included the 123bp region with *CHD-1* identity (fig. 4). Figure 12 shows that in the chicken hybridization was with a fragment of 3.1kb in the female only and with fragments of 1.5 and 6.0kb in both sexes. In the gull hybridization is similarily with a female specific fragment of 4.0kb a fragment of 3.0kb in males and females.

Delmas *et al.*, (1993) have already demonstrated the universal occurence of the *CHD-1* in the mammals. The evidence this blot provides, which features species representing both the major divisions of the birds, suggests that a minimum of two types of CHD gene exist in this Class. The first we termed *CHD-W* to denote its W linkage. The 123bp region from the great tit would appear to be a short exon from this gene. The second hypothetical gene is closely related to *CHD-W* and we have it termed *CHD-1A*, where the A denotes its avian nature. This gene is either Z or autosomally linked as it occurs in both sexes.

**Isolation of *CHD-1A***

The SacI/HindIII great tit probe was used at low stringency to screen a lZap II cDNA library from stage 10-12 (33-49hrs after the appearance of the primitive streak) chicken embryos. A plating of 2x$10^5$ plaques yielded a panel of 25 positive clones, 19 of these continued to hybridize intensely after purification. From three clones Z4, Z6 and Z11 a composite 6313 nucleotide sequence in fig. 5 was determined using the strategy illustrated in fig. 6.

The insert from the Z6 clone (bases 122-4132; fig. 5) and a BglII fragment of the Z4 clone (bases 3763-5007; fig. 5) were used separately to screen a similar number of plaques from a second cDNA library constructed from 10 day old chicken embryos. This screening identified a total of 45 positives of which 16 were found to have sequence identity with the composite sequence derived from the first library. Two additional clones contained a closely related sequence that is dealt with below.

A proportion of the clones from both libraries show variation from the sequence given in fig. 5 in one respect. Clones Z1, Z13, Z17, Z20 and Z23 are identical to the composite sequence 5' to base 4333 from there they terminate in an additional 37 to 163 bases of a new sequence that is identical in all five. Two clones from the second library CC43 and CC56 have 22 or 254bp of the same sequence at their 5' ends. Downstream of this motif both clones regained homology with the composite sequence at base 4334 and show no further deviation from the original sequence. From these seven clones a composite 264bp sequence can be derived and this is illustrated in fig. 7. None of the seven clones contain the whole of this sequence. Moreover, none of the ten clones that span the 4333/4334 insertion point contain this additional region. If inserted at this position the motif has an in frame, open reading frame spanning its entire length. The motif is extremely adenosine rich and this makes the amino acid lysine extremely common in the putative translation (see fig. 7). There are no splice donor or acceptor sites within the motif suggesting it is a final rather than an intermediary product of splicing.

Hybridization of a probe running from 2238 to 4132bp of the sequence chicken sequence to a blot of *PvuII* cut, male and female chicken genomic DNA shows that hybridization occurs to fragments that are both W and autosomally or Z chromosomally located. The level of hybridization is significantly stronger to the fragments common to both sexes suggesting that the probe represent the *CHD-1A* gene.

*CHD-1A* is very closely related to the mouse *CHD-1* gene being 79.8% identical in a 5035nt overlap. At the amino acid level the identity is raised to 90% over 1750 residues. In comparison to the mouse sequence we have yet to obtain the most 68bp 5' of the coding region. We do have an additional 1202bp of the 3' untranslated region but have not encountered a clone with an AATAAA termination signal or a 3' homopolymeric T tail. Both mouse and chicken sequences contain a stop codon in the same relative postions and sequence similarity is insignificant after this point. The published mouse sequence does not contain the additional 264bp motif described above.

The database search also identified an unpublished chicken derived sequence tagged as a delta crystallin binding protein (*DCBP*), with even greater identity than the mouse *CHD-1* gene: 99% over 2293 bp and 94% over 571 amino acid residues. The *DCBP* sequence is of 2292bp which extends over nucleotides 1922 to 4214 of *CHD-1A* (fig. 5). Despite the high nucleotide sequence identity the region of amino acid similarity does not extend the full length of the *DCBP*. This is due to apparent deletions in the DCBP clone that provides an initiation

methionine codon (257nt *DCBP*) and a stop codon (1939nt *DCBP*). The extremely high sequence identity, the fact that identity is maintained after the the apparent stop in the *DCBP* sequence, that none of the 41 CHD-related clones we found have exact sequence identity and that only small sequencing mistakes would be required to introduce false stop and start codons suggests that the DCBP sequence is *CHD-1A* but has been sequenced slightly inaccurately. Further evidence is required to confirm this however.

The database search with the whole *CHD-1A* gene also revealed significant identity to a previously unidentified portion of a 15 kb region of *S. cerivisiae* chromosome V. This region comprises an open reading frame of 4.4kb which lies between the *RAD4* [Gietz, 1988 #553] and the poly-A binding protein [Sachs, 1986 #554] gene coding regions. In an overlap of 1538 amino acids, the whole of the yeast open reading frame, there is an identity of 37.7% and a similarity of 59% (fig. 10) . The degree of conservation this similarity implies suggests the yeast sequence encodes a homologue of *CHD-1A* that we shall term *CHD-1Y* for the sake of discussion.

Delmas *et al.*, (1993) identified four motifs in *CHD-1* with possible functional significance. *CHD-1A* retains such close homology to *CHD-1* that these regions are virtually unchanged and are likely to perform similar functions as they do in the mouse.

The first motif is a chromodomain [Paro, 1991 #457] which falls between residues 274 and 311 (fig. 9). Figure 11 compares the the amino sequence of this region to that of eight others identified through a search of the EMBL database. The sequences fall into three categories. The first comprises the domain from *CHD-1*, *CHD-1A* and *CHD-1Y*. The second and third chromobox groups have been previously identified by [Pearce, 1992 #424]. The HP1 class comprises the *Drosophila* [James, 1986 #556] and human [Saunders, 1993 #555] *HP1* genes and two murine modifier (*Mod*) genes [Singh, 1991 #420]. The HP1 class is characterized mainly by glutamic acid rich block of six residues upstream of the chromobox. The third group, the Pc class, comprises the *Drosophila Pc* gene [Paro, 1991 #457] itself and its putative murine homologue the *Mod3* gene [Pearce, 1992 #424].

A search of the EMBL data base with the *CHD-1A* putative helicase domain (residues 451-911, fig. 9) raises the identity between this and *CHD-1Y* to 55% in an overlap of 471 amino acids. There is also significant, but lesser identity to, the putative helicase motifs in the human [Okabe, 1992 #558], and *S. cerivisiae* [Laurent, 1992 #454] *SNF2* gene, human [Muchardt, 1993 #557] and *Drosophila Brahma* [Tamkun, 1992 #458], *S. cerivisiae NPS1/STH1*[Laurent, 1992 #454; Tsuchiya, 1992 #560], human excission repair protein *ECCR6* [Troelstra, 1992 #559] and the *RAD54* [Emery, 1991 #562] and *MOT1* [Davis, 1992 #461] genes of *S. cerivisiae*. It should be noted that none of these latter genes contain a chromobox.

Only the four CHD genes show significant homology to the the third motif, a DNA binding region identified by Delmas *et al.*,(1993), whilst only *CHD-1A* and *CHD-1* have the three short basic HSDHR motif near the carboxy terminus, although this region is yet to be sequenced in

*CHD-W*. The *CHD-1Y* gene apparently terminates before this point so does not share th motif.

**Isolation f *CHD-W***

Two, CC14 and CC4, of eight *CHD-1* related clones isolated from the 10 day chick embyro library using 122-4132nt of *CHD-1A* as a probe, overlap (fig. 5) to provide the 1316bp of sequence given in fig. 8. This is a sequence closely related to, but distinct from *CHD-1A*. Identity over the 1316bp overlap is 90.5% and 90.1% at the nucleotide and amino acid level respectively. An alignment of the putative translations of *CHD-1*, *CHD-1A* and *CHD-W* is given in fig. 9. The amino acid identity between *CHD-1* and *CHD-1A* at 93.4% is marginally lower than that between that of *CHD-1* and *CHD-W*, 94.2%, over the same region

The 1335bp insert of CC4 was used at moderate stringency to probe a male/female, *PvuII* cut genomic blot featuring mouse, ostrich (*Struthio camelus*), chicken, bee-eater and hyacinth macaw (*Anodorhynchus hyacinthinus*; fig. 13). Hybridization with the mouse and ostrich shows no evidence of any sex linkage, bands of the same size and equal intensity appearing in both sexes. Hybridization with the ostrich is paricularly strong, greater even than with the cognate sequence in the chicken. This suggests that the genome size of the ostrich is considerably smaller than that of the chicken.

In all the bird species apart from the ostrich hybridization occurs with two types of fragment some that are female unique and others that are shared between the sexes. In the chicken some of the latter are of the same size as those hybridizing with the *CHD-1A* probe and result from cross hybrization under the conditions of low stringecy that we employed. When probed with the CC4 sequence it is clear that hybridization with the female linked fragments is far stronger, at least in the chicken than with the shared fragments (bear in mind, also, that the female chicken only has a single dosage of the W linked gene). This indicates that CC4 is W linked and represents part of *CHD-W*.

There is also some indication that the intensity of hybridization to the male shared bands is stronger than that to the corresponding female fragments. This suggests that the *CHD-1A* gene is present in different dosages in the two sexes which would result from it being located on the Z chromosome. This remains to be confirmed.

## Discussion

The female specific great tit probe GT-W was described by Griffiths and Tiwari (1993) as a means of identifying sex in this species. The results presented here suggest this sequence represents part of a intron in a W linked gene. By moving downstream from this sequence it

has been possible to isolate an putative exon from a gene that we have named *CHD-W* due to its close sequence identity to the mouse *CHD-1* gene [Delmas, 1993 #415] and its W location.

Using the *CHD-W* fragment we attempted to isolate a similar, W linked sequence that Southern blot analysis had shown was present in the chicken. From several clones a 6313bp cDNA sequence was assembled but although it has close sequence identity to the great tit *CHD-W* fragment Southern blot analysis shows it is not located on the W chromosome. This second gene was termed *CHD-1A* (A = avian). This blot shows a second gene closely related to *CHD-1A* is W located. This sequence could not be cloned from a stage 10-12 chick cDNA library although 19 *CHD-1A* clones were isolated. However, two clones yielding ? bp of a second CHD gene were isolated along with a further 14 *CHD-1A* clones from a day 10 chick cDNA library. Southern blot analysis showed that this second clone was W chromosome derived and so represents *CHD-W*. Attempts are underway to isolate the remainder of *CHD-W*.

Southern blots of a variety of bird species showed that *CHD-W* is W chromosome linked in all birds except the ostrich. This suggests that the gene is sex linked throughout the class with the exception of the primitive ratites which the ostrich represents where it appears to be autosomally located.

An alternative explanation is that the *CHD-W* is in fact W linked in ratites but occurs in a region of the W chromosome which still recombines with the Z chromosome. If *CHD-1A* were Z linked, then recombination between Z and W linked copies of CHD would maintain their sequence identity resulting in the apparently autosomal location indicated by the Southern blot. A mammalian example would be the *MIC2* and *STS* genes that are located in the pseudoautosomal region of the Y chromosome [Ellis, 1989 #563] and would give analogous results to those observed here.

Two lines of evidence support this alternative hypothesis. The first is that the Southern blot analysis suggests that *CHD-1A* is Z linked in non-ratites which would make the chromosomal location of the CHD-genes consistant throughout the class. Hybridization of *CHD-1A* to genomic blots is apparently stronger to fragments from male birds which would result from this sex having two copies of any Z linked gene in comparison to a single copy in the female (this result is not clear cut and requires confirmation by chromosomal *in situ*). The second line of evidence is that the sex chromosomes of the ratites are not morphologically differentiated as is the case with other birds [Christidis, 1990 #440]. Morphological similarity suggests recombination still occurs between extensive regions of the ratite Z and W which may include the CHD genes and so produce the pattern of hybridization observed.

Although we have yet to clone the whole of *CHD-1A* the 6313bp sequenced so far shows a close identity to the mouse *CHD-1* gene over the putative coding region. It also includes all four features identified by Delmas *et al.* (1993) as having possible functional significance. This includes a chromodomain, a helicase, a DNA binding motif and a basic, five amino acid motif that is repeated three times (Fig. 9). The similarity of the sequence derived thus far from *CHD-*

*W* to that of *CHD-1* and *CHD-1A* suggest it will be of similar length and possess these sa motifs. We have also identified an alternatively spliced form of *CHD-1A* and *CHD-W* wh has a similar adenine rich motif inserted at an identical point (nt 4033-4034 *CHD-1A* and nt ? *CHD-W*). The exact form of these alternative mRNAs is yet been elucidated. It is interesting to note that we obtained no clones that spanned these breakpoints which contained this additional motif; the sequence was built up from partial sequences derived from either 5' or 3' terminii of different clones. Delmas *et al.*, (1993) produced a mRNA Northern blot probed with fragments of *CHD-1* occurring 5' to this breakpoint and discovered an mRNA species of about 4kb. This would correspond to a species cleaved near this insertion point. What purpose this would serve is unknown. Moreover the putative yeast homologue of *CHD*, *CHD-1Y*, which was identified from amino acid identity to *CHD-1A* from the genomic sequence on the EMBL database does not apparently have a similar motif. This is suggested because the *CHD-1Y* sequence was derived from a genomic clone which would allow the identification of any such sequence were it to be spliced in in the normal manner.

The significance of the four functional domains found in the CHD genes will be discussed in turn. The first, the carboxy-terminal trimer repeat of five basic amino acid residues, has no known function and is not shared by any other sequences from the EMBL database. Furthermore, the *CHD-1Y* gene which is truncated by a little over 200 amino acid residues in comparison to *CHD-1* and *CHD-1A* does not contain this motif.

The second functional domain was identified by Delmas *et al.* (1993) as having sequence selective DNA binding capacity. Whether this is highly specific or just to A+T rich regions was not established. They also noted that this domain contains Lys-Arg-Pro-Lys-Lys and Arg-Gly-Arg-Pro-Arg motifs which enable genes like *HMG-1*, *D1* and *Engrailed* to bind in the minor groove of A+T rich DNA.

A third functional motif is located towards the N-terminus of the CHD-protein and is termed the chromodomain [*Chr*omatin *O*rganization *Mo*difier; Paro, 1990 #459]. This is a highly conserved domain of between 37-50 amino acids that has been shown to be represented in the genomes of plants, nematodes, insects and vertebrates [Singh, 1991 #420]. Several chromobox genes have been isolated from human, mouse and *Drosophila* and have been divided into the polycomb (Pc) class and the heterochromatin protein-1 (HP1) class on the basis of related structure [Pearce, 1992 #424]). The CHD-genes have a distinct form of the choromobox characterized by close homology between yeast and vertebrate forms in the 5' half of the box itself but extending a further 17 residues downstream. These differences indicate that this form of the chromobox defines a third subgroup the CHD class

The *Pc* gene forms one of a eponymously named group (Pc-g) of about 12 genes defined through homeotic mutants in *Drosophila* that prevent fixation and maintainance of a determined state. They act as transcriptional repressors of homeotic genes, notably of the antennapedia complex [ANT-C; Paro, 1990 #459]. Members of the ANT-C and the other major group of

Drosophila homeotic genes, the bithorax complex (BX-C), are responsible for defining segmental identity during development [Lewis, 1978 #465; Kaufman, 1980 #464]. Initially, their expression patterns are designated by early acting maternal and segmentation genes (see 4,6,7 kennison). However, these maternal genes are only transiently expressed. During the later stages of development their role as transcriptional activators is adopted by an assemblage of genes including the trithorax group (Trx-g), whilst many of their repressive effects are assumed by the Pc-g [Kennison, 1993 #416].

The *polycomb* (*Pc*) gene itself is perhaps the best studied member of the Pc-g. Zink and Paro (1989) used *Pc-B*-galactose fusion proteins to show that it binds to around 100 different sites on the polytene chromosome including loci where other members of the Pc-g are located. Any disruption of the chromodomain abolishes the specificity of this reaction [Messmer, 1992 #452]. However, the Pc-g protein appears to lack any type of endogenous DNA binding capacity so it is thought that it acts as part of a protein complex with other components that are responsible for the site specific DNA binding [Paro, 1990 #459].

The repressive effects of the Pc-g are are thought to be the result of chromatin compaction. In other words, the DNA is packaged into heterochromatin to prevent or reduce the expression of functional genes [Paro, 1990 #459]. This is a mechanism related to position effect variegation (PEV; [Henikoff, 1990 #460]), to dosage compensation in mammals which sees the complete heterochromatization of one of the female's X chromosomes and possibly to gene imprinting whereby the expression of maternally and paternally inherited alleles differs [Peterson, 1993 #565]. The links with PEV have recently been substantiated in that HP1, a recognized modifier of PEV, and *Pc* both contain chromodomains [Paro, 1991 #457]. Like the *Pc* protein, HP1 appears to form part of a structural complex that transforms euchromatin to heterochromatin. Furthermore, both PEV and the repressive effects of *Pc* are passed, in a clonal manner, to daughter cells ([Henikoff, 1990 #460; Struhl, 1981 #466]; a characteristic also of gene imprinting.

With the CHD-type gene containing both a DNA binding motif and a chromobox it may appear reasonable to suggest that they encode repressors with an endogenous, site selective DNA binding system. However, CHD genes contain a further functional motif that is structurally related to the helicases. The sequence identity is closest to the yeast *SNF2/SWI2* [Abrams, 1986 #467] and Drosophila *Brahma* genes [Tamkun, 1992 #458], both of which are transcriptional *activators*. Indeed, *Brahma* is part of the Trx-g which are considered direct antagonists to the Pc-g. Other genes which contain more distantly related helicase domains are involved in DNA repair and chromatid separation during mitosis [Laurent, 1993 #453; Sung, 1993 #468].

The *SWI2* gene product has been shown to enhance the transcription of other genes probably as part of a complex that includes *SWI1*, *SWI3*, *SNF5*, *SNF6* and in conjunction with

gene specific DNA binding proteins [Laurent, 1991 #469; Peterson, 1992 #456]. A mode of action strikingly similar to that of *Pc*.

Although it remains to be formally demonstrated that *SWI2* is a helicase, it does have close structural similarities with proven helicase genes and also possesses the required DNA stimulated ATPase activity [Laurent, 1993 #453]. Laurent *et al.*, go on to postulate that the *SWI2* containing complex may act by two mechanisms acting either separately or in conjunction. In the first they envisage helicase mediated DNA melting to allow the egress of RNA polymerase II. Alternatively *SWI2* could allow chromatin remodelling, in effect overcoming any inhibitory packaging of the DNA and so enhancing transcription.

The juxtaposition of a helicase and a chromodomain within the same gene presents a paradox that may challenge the perceived roles of the two motifs. A simple explanation is that alternative splicing could remove one or other of these domains prior to translation. However, there is little support for this idea from the work of ourselves or Delmas *et al.*, (1993).

An alternative explanation could be due to our lack of real knowledge about the function of the chromobox. Whilst it is well established that helicases do disassociate DNA and so facilitate transcription [Matson, 1990 #470], the role of the chromodomain in repression is based on more circumstantial evidence. *Pc*, as we have seen, does not bind DNA itself although mutations in the chromobox prevent the formation of site specific complexes. It is possible that the chromodomain is involved more in maintaining the structural integrity of the repressive complex than in the repressive mechanism itself. Based on this supposition, the CHD-protein may form a different type of complex able to bind at a site dictated or influenced by its own binding domain and activate these loci via helicase activity.

While both this scenario is speculative it is probable that CHD-type genes are active during development and are able to bring about heritable changes in transcription. The presence of an endogenous DNA binding domain suggests it has fewer targets than *Pc*, for example, which could form part of several different active complexes. With *CHD-W* being confined to the W chromosome is likely to have a role in some aspect of female development and we suggest this may be critical to the determination of gender. In support this hypothesis we were unable to find any *CHD-W* clones in a library constructed prior to sex determination which occurs at day 7 [Lutz-Ostertag, 1954 #564] but were able to isolate two clones from a smaller pool of candidates at day 10. This suggests that the expresion of *CHD-Y* may occur at a time consistant with it having a sex determining role.

If *CHD-W* alone or in conjunction with *CHD-1A* cauases sex determination in birds then then several potential mechanisms are plausible.

CHD-W ?

(1) In the simplest scenario *CHD-Y* may act as a simple trigger like *SRY* [Koopman, 1993 #550] to either cause expression or repression of downstream genes in order initiate testis development.

(2) *CHD-W* may interact with other autosomal or Z linked genes whereby the dosage of *CHD-W* in comparison these other factors causes initiates development down the male or female pathways.

A more complicated scenario is if *CHD-W* acts in together with *CHD-1A* to cause sexual differentiation. Different mechanisms could operate depending whether *CHD-1A* turns out to be Z linked as we suspect or autosomal.

(3) If *CHD-1A* is Z linked, then male birds get two doses of the *CHD-1A* expression product to one in female birds. Perhaps the 1:1 ratio of functionally distinct *CHD-1A* and *CHD-W* products is what initiates female development whilst a double dosage of *CHD-1A* results in males.

(4) Alternatively, just the single dosage of Z linked *CHD-1A* product could result in female development and expression of *CHD-W* only occurs after sexual differentiation to equalize dosages of functionally similar proteins.

(5) If *CHD-1A* is autosomal however, it could be envisaged that *CHD-1A* and *CHD-W* are functional homologues and the three doses in females (AAW) is required to promote female development, whilst the double dosage in males (AA) causes the differentiation of the testis and the development of the male phenoype.

The evidence from aneuploid chickens discussed in the introduction, does suggest that the mechanism that does operate involves some degree of dosage dependence which tends to exclude mechanism (1). However the similariy of *CHD-W* to *HP1*, the *Pc* protein and other transcriptional modifiers that act through chromatin remodelling show that the expression of this type is crucially dependent on dosage [Locke, 1988 #471] . With the different dosages of gene product and/or potential target sites that aneuploids possess it may be that anlysis of these type of mutants has, thus far, served to confuse the issue.

**Claims**

1. The nucleotide sequences of *CHD-1A* and *CHD-W* as shown in fig. 5, fig. 7 and fig. 8.
2. A clone or subclones of *CHD-1A* and *CHD-W* as defined in 1.
3. A fragment of *CHD-1A* and *CHD-W* capable of giving W specific signal on hybridization to a non-ratite bird.
4. A fragment of *CHD-1A* and *CHD-W* obtainable by restriction endonuclease digestion therof and being capable of giving a W specific signal on hybridization to genomic DNA of a non-ratite bird.
5. A clone or subclone of a fragment according to either of claims 3 and 4.
6. A nucleic acid or fragment or oligonucleotide having substantially the sequence of *CHD-1A* and *CHD-W* as set out in fig. 5, fig. 7 and fig 8.
7. A clone or a subclone of a nucleic acid or fragment according to claim 6.
8. A nucleic acid or fragment or oligonucleotide having susbtantially the same sequence of the chicken or great tit CHD-gene as set out in figs 1, 3, 5, 7 or 8.
9. A nucleic acid or fragment or oligonucleotide being capable of giving a W chromosome specific signal on hybridization to the genomic DNA of a non-ratite bird.
10. A nucleic acid or fragment or oligonucleotide according to claim 4 or claim 9 capable of giving W chromosome specific signal on hybridization to the genomic DNA of a chicken, turkey, duck, parrot.
11. A nucleic acid or fragment or oligonucleotide according to any one of claims 4, 9 and 10 capable of giving W chromosome specific signal on hybridization to the genomic DNA of a non-ratite bird under conditions of high stringency.
12. A nucleic acid or fragment or oligonucleotide according to any one of claims 4, 9 and 10 capable of giving W chromosome specific signal on hybridization to the genomic DNA of a non-ratite bird under conditions of low stringency.
13. A nucleic acid or fragment or oligonucleotide according to any one of the claims 9 to 13 containing substantially the sequence of the chicken CHD-gene as set out in fig. 5, fig. 7 and fig. 8.
14. A nucleic acid or fragment or oligonucleotide encoding a CHD-protein, fragment thereof or polypeptide containing a CHD-gene or part thereof or encoding a CHD-mimetope protein or fragment thereof or CHD-mimetope polypeptide.
15. A process for ascertaining the sex of an embryo, foetus, cell, tissue or organism comprising hybridizing a nucleic acid of fragment or oligonucleotide according to any one of claims 1 to 14 with DNA or RNA of the embryo, foetus, cell, tissue or organism or with cDNA reverse transcribed from RNA of the embryo, foetus, cell, tissue or organism or with cDNA or DNA amplified by cloning or polymerase chain reaction from DNA or RNA of the embryo, foetus, cell, tissue or organism.

16. Use of a nucleic acid or fragment or oligonucleotide of any one of claims 1 to 14 in ascertaining the sex of an embryo, foetus, cell, tissue or organism.
17. A process for controlling the sex of the progeny of an organism comprising inserting a nucleic acid or fragment or oligonucleotide of any one of claims 1-14 into the genome of the organism or progenitor thereof.
18. Use of a nucleic acid or fragment or oligonucleotide of any one of claims 1 to 14 in controlling the sex of the progeny of an organism.
19. A CHD-protein, fragment thereof or polypeptide containing a CHD-gene of part thereof or a CHD-mimetope protein, fragment thereof or a CHD-mimetope polypeptide.
20. A protein or fragment thereof or polypeptide containing a CHD-chromobox including at least one of the characteristic amino acid residues at position 11, 12, 20, 27 or 34 inside the chromobox or 3, 6, 8, 12-15 or 16 directly downstream of the chromobox when aligned to best effect and as set out in fig. 11.
21. A protein or fragment thereof or a polypeptide encoded by a nucleic acid or fragment or oligonucleotide according to claims 1-14 and containing a CHD-chromobox
22. A process for controlling the sex of the progeny of an organism comprising supplyong exogenously to a cell of the organism or a progenitor of the organism a protein or fragment thereof or a polypeptide according to any one of claims 19-21
23. A process according to claim 22 wherin the protein or fragment thereof or polypeptid is supplied and activates a *CHD-1A* or *CHD-W* target gene.
24. An antibody or fragment thereof against a protein or fragment thereof or polypeptide according to any one of claims 19-21.
25. An antibody producing cell capable of expressing an antibody or fragment thereof according to claim 24.
26. Use of a protein or fragment thereof or polypeptide according to any one of claims 19-21 or antibody or fragment thereof or cell according to claims 24 or 25 in ascertaining the sex of an embryo cell tissue or organism.
27. A transgenic or chimeric animal having a heterologous nucleic acid or fragment or oligonucleotide according to any one of claims 1 to 14 in the genome of at least the germ cells of the animal.
28. Gametes of an animal according to claim 27.
29. Progeny of an animal according to claim 27.
30. Progeny according to claim 29 which are transgenic or chimeric and have a heterologous nucleic acid or fragment according to any one of claims 1-14 in the genome of at least the germ cells of the progeny.
31. A method of controlling the population of a species of bird which comprises introducing an individual member of the species into the population, said individual having a copy or copies

of a nucleic acid fragment or oligonucleotide according to any one of claims 1 to integrated on a chromosome (carrier chromosome) be it sex linked or autosomal whereby when the male breeds with other individuals of the population the progeny are sustantially of one sex or are sexually dysfunctional intersexes.

32. A method according to claim 31 where the nucleic acid integrated into the carrier chromosome is homologous to the native *CHD-1A* or *CHD-W* gene of the bird.

ble 1. Sex of domestic fowl with normal and abnormal chromosome complements (from IcCarrey, 1979 #448] and [Crew, 1954 #450])

| Chromosome complement | Phenotype |
|---|---|
| AA ZZ | Male |
| AA ZW | Female |
| AA ZZW | Male ? |
| AA ZZZ | Male |
| AAA ZZZ | Male |
| AAA ZZW | Intersex/male |

Figure 1. The DNA sequence of the pGT-W insert

```
CCCGGTCGGAGGTTTCAAGGAATGACTAGATGTGGCACTTAGTGCCATGGTCTAGTTGAC      60
AAGGTGATGGTTGGTCAAAAGTTGGACTCGATGATCTCAGAGTTTTTTTCCAGCCTTAAT     120
AATTCTATGAATTCTGTAATTTTATTCTTGATCTTTTTGAGCGAAGTTTGTTTGGGGATT     180
TTAGTTTGGTTTCCCTGTCACTGTTTTCTTTCCTTGAAACTGACTTTCATTTGCAACATG     240
AGAATTGCTGTATTTGTCAGGTTACAAGTAGTGCAATGGCTGCTTAGAAGTAGTGAGAAA     300
CATTTAGGGAAATACTGGAGTGAAGCAAACACAGTGGTACTGCCAAACTGTAGCTTTGGG     360
ATTTGAGGAGCCACAGAGTTGTATATAAATTTGTTTAATGATATCCTGCCCCTGCCTTCC     420
ATTAATTGCTTGTTTTATGAAACCACTCTTTTTTTTTTTTTTTTTTTTTTGGCTTCTTCA     480
TATCCTGTGGTAATGAGTTAATGCATTTAGAAGCACATGGCAGAACTAGGAGATCTGTGG     540
ATGACAGTGGTACAGGAGCTCTGAATTTTTTAGATAAACTATGAGAGTGGAAACAGAAAT     600
CTGAGGCTAGTTTCTTGAGCTGACTGTAAATTTTGTGAGAATATTTTCAAGACTACATTA     660
GTTGTGTGTTTGAGGAAAAATAAAATGTTTAAGTTGTCCATTCCTTGAAACCTCCCGACC     720
GGG                                                               723
```

Figure 2. A map of the 9.6kb insert of the lFixII clone isolated from the great tit using pGT-W. pGT1.7 and pGT8 are the two *EcoRI* subclones into which the fragment was divided. The broken line corresponds to the region with absolute sequence identity to the pGT-W insert. The position of the region with identity to the mouse *CHD-1* gene is indicated.

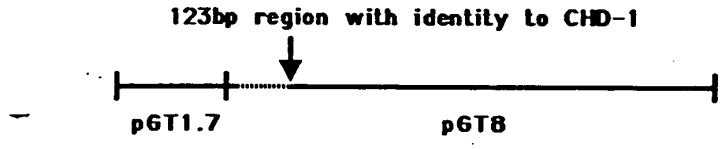


Figure 3. An alignment of 123bp fragment of the great tit (GT) *CHD-W* gene in pGT8 with the autosomal/Z located chicken (C) *CHD-1A* the chicken *CHD-W* gene and bases 3855-3977 of the mouse (M) *CHD-1* gene. An alignment of the deduced amino acid sequence is also given.

```
M  CHD-1    ATTCTTCCAG ATGATCCTGA TAAAAAACCA CAAGCAAAAC AGTTACAGAC
C  CHD-1A   ATTTTACCTG ATGATCCAGA CAAGAAACCC CAGGCAAAGC AGCTACAGAC
C  CHD-W    ATTTTACCTG ATGATCCAGA TAAGAAACCC CAGGCTAAGC AGTTACAGAC
GT CHD-W   ATTTTACCTG ATGACCCAGA TAAGAAACCA CAGGCAAAGC AGTTGCAGAC

M  CHD-1    CAAAAAACCA CAAGCAAAAC AGTTACAGAC CCGTGCAGAC TACCTCATCA
C  CHD-1A   CAAGAAACCC CAGGCAAAGC AGCTACAGAC CCGTGCAGAC TACCTCATTA
C  CHD-W    CAAGAAACCC CAGGCTAAGC AGTTACAGAC CCGTGCAGAT TACCTCATTA
GT CHD-W   CAAGAAACCA CAGGCAAAGC AGTTGCAGAC CCGTGCAGAT TACCTCATTA

M  CHD-1    AACTACTTAG CAGAGATCTT GCAAAAAGAG AGGCTCAGAG ACTTTGTGGT GCG
C  CHD-1A   AATTACTGAA TAAAGACCTT GCAAGAAAGG AAGCACAAAG GCTTGCTGGT GCA
C  CHD-W    AATTACTGAA TAAAGACCTT GCAAGAAAGG AAGCACAGAG ACTTGCTGGT GCA
GT CHD-W   AATTACTGAA TAAAGACCTT GCAAGAAAAG AAGTGCAAAG ACTTACTGGT GCA
```

```
M  CHD-1          ILPDDPDKKPQAKQLQTRADYLIKLLSRDLAKREAQRLCGA
C  CHD-1A         ILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEAQRLAGA
C  CHD-W          ILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEAQRLAGA
GT CHD-W          ILPDDPDKKPQAKQLQTRADYLIKLLNKDLARKEVQRLTGA
                  **************************** **** * *** **
```

Figure 4. The section of pGT8 that hybridized to a female specific fragment of 3.1kb in the chicken. This probe was also used to screen the chicken cDNA library. The hatched line represents the female specific great tit motif shown in fig. 3.

Figure 5. A partial nucleotide sequence of *CHD-1A* as defined by the clones Z4, Z6 and Z11. Two asterisks mark the position where part of the sequence illustrated in fig. 7 is spliced onto the 5' or 3' ends of a proportion of the clones isolated.

```
GGCTGGGTCA GCTTCAGGTT CTGGATCTGG TTCAAGCTCT GGAAGCAGTA GCGATGGAAG TAGCAGCCAG TCAGGTAGCA
        10         20         30         40         50         60         70         80
GTGACTCTGA ATCTGGTTCA GAGTCAGGCA GTCAATCCGA ATCAGAGTCT GACACATCTA GAGAGAAGAA ACAAGTTCAA
        90        100        110        120        130        140        150        160
GCTAAACCTC CGAAAGCTGA CGGATCTGAG TTTTGGAAGT CCAGTCCAAG CATACTTGCT GTACAGAGAT CAGCAGTGCT
       170        180        190        200        210        220        230        240
CAAGAAGCAA CAGCAACAGC AAAAAGCAGC ATCATCAGAC AGTGGTTCAG AAGAGGACTC ATCCAGTAGT GAAGATTCTG
       250        260        270        280        290        300        310        320
CCGATGATTC GTCCAGTGAA ACTAAGAAGA AAAAGCATAA AGATGAAGAC TGGCAAATGT CAGGGTCAGG GTCAGTATCA
       330        340        350        360        370        380        390        400
GGAACTGGTT CTGATTCTGA ATCGGCGGAA GATGGGGATA AAAGCAGTTG TGAAGAAAGT GAATCTGACT ATGAGCCAAA
       410        420        430        440        450        460        470        480
AAACAAAGTC AAAAGCCGTA AACCTCCAAG CAGAATTAAG CCAAAAAGTG GGAAAAAGAG CACAGGACAG AAGAAGAGGC
       490        500        510        520        530        540        550        560
AACTTGATTC ATCAGAGGAG GAGGAGGACG ATGATGAAGA TTATGATAAG AGAGGATCTC GTCGCCAGGC AACAGTGAAT
       570        580        590        600        610        620        630        640
GTTAGTTACA AAGAAGCTGA AGAAACCAAG ACAGATTCTG ATGATTTGCT GGAAGTTTGT GGAGAGGATG TCCCACAGAC
       650        660        670        680        690        700        710        720
TGAAGAAGAT GAATTTGAAA CTATAGAGAA GTTTATGGAC AGTCGAATTG GCCGAAAAGG AGCCACTGGT GCCTCAACCA
       730        740        750        760        770        780        790        800
CCATCTATGC CGTTGAGGCA GATGGTGACC CAAATGCTGG GTTTGAAAAG TCAAAGGAGC TGGGAGAAAT ACAGTATCTT
       810        820        830        840        850        860        870        880
ATTAAATGGA AAGGCTGGTC ACACATCCAT AACACTTGGG AAACTGAAGA AACGCTGAAG CAACAAAATG TTAAAGGAAT
       890        900        910        920        930        940        950        960
GAACAAACTG GACAACTACA AGAAAAAGGA TCAGGAGACA AAACGCTGGC TGAAAAATGC TTCTCCAGAA GATGTGGAAT
       970        980        990       1000       1010       1020       1030       1040
ATTATAACTG CCAGCAGGAG CTTACAGATG ATCTGCACAA ACAATATCAA ATAGTGGAAA GAATAATTGC TCATTCAAAT
      1050       1060       1070       1080       1090       1100       1110       1120
CAAAAGTCAG CAGCTGGTTA TCCGGACTAC TATTGCAAAT GGCAGGGTCT GCCTTACTCA GAATGTAGCT GGGAAGATGG
      1130       1140       1150       1160       1170       1180       1190       1200
TGCTCTCATT GCCAAAAAGT TTCAGGCACG CATTGATGAG TATTTTAGCA GAAATCAATC CAAGACTACT CCCTTTAAGG
      1210       1220       1230       1240       1250       1260       1270       1280
ACTGCAAGGT TCTAAAACAG AGACCAAGAT TTGTTGCACT GAAGAAGCAA CCATCTTACA TTGGAGGACA TGAAAGTCTG
      1290       1300       1310       1320       1330       1340       1350       1360
GAGTTAAGAG ATTATCAGTT AAATGGATTG AATTGGCTCG CTCATTCATG GTGCAAAGGA AATAGTTGTA TTCTTGCAGA
      1370       1380       1390       1400       1410       1420       1430       1440
TGAAATGGGT CTGGGTAAAA CAATACAAAC AATTTCTTTT CTGAACTACC TGTTTCATGA ACATCAACTG TATGGCCCTT
      1450       1460       1470       1480       1490       1500       1510       1520
TTCTTCTGCG CGTGCCACTT TCTACCTTGA CATCTTGGCA AAGAGAGATT CAAACTTGGG CTCCTCAGAT GAATGCTGTA
      1530       1540       1550       1560       1570       1580       1590       1600
GTTTACTTAG GAGATATAAC TAGTAGAAAT ATGATAAGGA CTCATGAATG GATGCATCCA CAGACTAAAC GATTAAAGTT
      1610       1620       1630       1640       1650       1660       1670       1680
TAACATACTT CTGACGACAT ATGAAATTTT ACTGAAGGAT AAGTCATTCC TTGGTGGTCT CAATTGGGCA TTCATAGGAG
      1690       1700       1710       1720       1730       1740       1750       1760
TTGATGAAGC TCATCGTTTA AAAAATGATG ACTCTCTTCT GTACAGGACT TTAATAGACT TTAAGTCCAA CCATCGACTT
      1770       1780       1790       1800       1810       1820       1830       1840
CTGATTACTG GAACCCCACT GCAAAATTCC CTCAAAGAGC TGTGGTCTTT GTTGCATTTC ATCATGCCAG AAAAATTTTC
      1850       1860       1870       1880       1890       1900       1910       1920
CTCCTGGGAA GATTTTGAAG AGGAGCATGG CAAAGGAAGA GAGTATGGTT ATGCAAGTCT TCACAAAGAG CTTGAACCAT
      1930       1940       1950       1960       1970       1980       1990       2000
TTTTACTAAG AAGAGTTAAA AAAGATGTAG AAAAGTCTTT ACCTGCTAAG GTTGAACAAA TTCTGAGGAT GGAAATGAGT
      2010       2020       2030       2040       2050       2060       2070       2080
GCATTGCAGA AGCAATATTA CAAGTGGATT TTAACAAGGA ATTATAAAGC CCTCAGTAAA GGTTCAAAAG GCAGTACCTC
      2090       2100       2110       2120       2130       2140       2150       2160
AGGCTTTCTG AACATTATGA TGGAACTTAA GAAGTGTTGT AACCATTGCT ACCTCATTAA GCCACCAGAT GATAATGAAT
      2170       2180       2190       2200       2210       2220       2230       2240
TCTATAATAA ACAGGAGGCC TTACAGCATT TGATACGTAG CAGCGGGAAA CTAATCCTTC TTGACAAGCT ACTGATTCGT
      2250       2260       2270       2280       2290       2300       2310       2320
CTGCGAGAAC GTGGCAACAG AGTTCTGATT TTCTCTCAGA TGGTGAGGAT GCTGGACATC CTAGCAGAAT ATCTGAAGTA
      2330       2340       2350       2360       2370       2380       2390       2400
TCGCCAGTTT CCCTTCCAGA GACTTGATGG ATCAATAAAA GGGGAATTGA GGAAGCAAGC ACTGGATCAT TTCAATGCAG
      2410       2420       2430       2440       2450       2460       2470       2480
AAGGATCAGA GGATTTCTGT TTTTTACTGT CTACAAGAGC TGGAGGATTA GGTATTAACT TGGCATCTGC TGACACTGTA
      2490       2500       2510       2520       2530       2540       2550       2560
GTTATTTTTG ATTCTGACTG GAATCCACAG AATGATCTGC AGGCACAGGC GAGAGCTCAT AGAATTGGAC AGAAGAAACA
      2570       2580       2590       2600       2610       2620       2630       2640
GGTTAATATT TATCGGCTAG TCACAAAAGG ATCAGTAGAA GAAGATATTC TTGAAAGAGC CAAGAAGAAG ATGGTGCTAG
      2650       2660       2670       2680       2690       2700       2710       2720
ACCATTTAGT AATTCAGAGA ATGGACACGA CAGGAAAAAC TGTTCTGCAT ACAGGTTCAA CTCCATCAAG CTCTACACCT
```

```
      2730        2740        2750        2760        2770        2780        2790        2800
TTAATAAAG AAGAGTTATC AGCTATTTTG AAGTTTGGTG CTGAGGAACT CTTTAAAGAA CCTGAAGGAG AAGAACAGGA
      2810        2820        2830        2840        2850        2860        2870        2880
CCCCAGGAA ATGGATATAG ATGAAATCTT GAAGAGAGCT GAAACTCGGG AAAATGAGCC AGGTCCATTG ACTGTAGGGG
      2890        2900        2910        2920        2930        2940        2950        2960
ATGAGTTGCT TTCACAGTTC AAGGTGGCGA ACTTTTCCAA TATGGATGAA GATGATATTG AGTTGGAACC AGAAAGAAAT
      2970        2980        2990        3000        3010        3020        3030        3040
TCAAGAAATT GGGAAGAAAT CATCCCAGAA TCCCAACGGA GAAGGATAGA GGAGGAGGAA AGACAAAAAG AACTTGAAGA
      3050        3060        3070        3080        3090        3100        3110        3120
AATATACATG CTCCCGAGGA TGAGAAACTG TGCAAAACAG ATCAGCTTTA ATGGGAGTGA AGGAAGACGC AGTAGGAGCA
      3130        3140        3150        3160        3170        3180        3190        3200
GAAGATATTC TGGATCTGAT AGTGACTCCA TCACAGAAAG AAAACGGCCA AAAAAGCGTG GAAGACCTCG AACCATTCCT
      3210        3220        3230        3240        3250        3260        3270        3280
CGAGAAAATA TTAAAGGATT TAGTGATGCA GAGATCAGGC GGTTTATCAA GAGTTACAAG AAATTTGGTG GCCCTCTGGA
      3290        3300        3310        3320        3330        3340        3350        3360
AAGGTTAGAT GCTGTAGCTA GAGATGCTGA ACTGGTTGAT AAATCTGAGA CAGACCTTAG ACGTTTGGGT GAACTTGTAC
      3370        3380        3390        3400        3410        3420        3430        3440
ATAATGGATG CATTAAGGCT TTAAAGGACA ATTCATCTGG ACAAGAAAGA GCAGGAGGTA GACTTGGGAA AGTTAAAGGC
      3450        3460        3470        3480        3490        3500        3510        3520
CCAACGTTTC GAATCTCAGG AGTGCAGGTG AATGCAAAAC TAGTCATCTC TCACGAAGAA GAGCTGGCAC CACTGCACAA
      3530        3540        3550        3560        3570        3580        3590        3600
ATCCATTCCT TCAGATCCAG AAGAAAGGAA AAGATATGTC ATCCCATGCC ACACCAAGGC TGCTCACTTC GATATAGATT
      3610        3620        3630        3640        3650        3660        3670        3680
GGGGTAAAGA AGATGATTCC AATCTGTTAG TAGGCATCTA TGAATATGGC TATGGCAGCT GGGAAATGAT AAAAATGGAT
      3690        3700        3710        3720        3730        3740        3750        3760
CCAGATCTCA GCTTAACACA GAAGATTTTA CCTGATGATC CAGACAAGAA ACCCCAGGCA AAGCAGCTAC AGACCCGTGC
      3770        3780        3790        3800        3810        3820        3830        3840
AGACTACCTC ATTAAATTAC TGAATAAAGA CCTTGCAAGA AAGGAAGCAC AAAGGCTTGC TGGTGCAGGC AATTCCAAGA
      3850        3860        3870        3880        3890        3900        3910        3920
GAAGGAAGAC AAGAAATAAG AAGAATAAGA TGAAGGCTTC AAAAATAAAA GAAGAAATAA AGAGTGATTC TTCACCACAA
      3930        3940        3950        3960        3970        3980        3990        4000
CCCTCAGAAA AATCTGATGA AGATGATGAG GAGGAGGATA ACAAGGTAAA TGAAATGAAA TCTGAAAATA AAGAAAAATC
      4010        4020        4030  **    4040        4050        4060        4070        4080
TAAAAAAATT CCATTGCTGG ATACTCCAGT TCATATTACT GCAACCAGTG AACCAGTTCC TATCTCAGAA GAATCTGAAG
      4090        4100        4110        4120        4130        4140        4150        4160
AACTCCATCA GAAGACATTT AGTGTGTGCA AAGAAAGAAT GAGGCCTGTC AAAGCAGCAC TGAAACAGCT GGATAGACCA
      4170        4180        4190        4200        4210        4220        4230        4240
GAGAAGGGCC TTTCTGAAAG GGAGCAGCTG GAACATACTA GGCAGTGTCT AATCAAAATT GGGGATCACA TTACAGAATG
      4250        4260        4270        4280        4290        4300        4310        4320
CCTGAAGGAG TACACAAATC CCGAGCAAAT AAAACAGTGG AGGAAAAATT TGTGGATTTT TGTGTCCAAG TTTACAGAAT
      4330        4340        4350        4360        4370        4380        4390        4400
TTGATGCCAG AAAGCTGCAC AAACTCTACA AACATGCAAT CAAAAAGCGC CAAGAGTCTC AGCAACACAA TGACCAAAAC
      4410        4420        4430        4440        4450        4460        4470        4480
ATTAGCAGCA ATGTGAATAC ACATGTAATC AGAAATCCAG ATGTGGAAAG ACTGAAGGAG ACTACAAACC ATGATGATAG
      4490        4500        4510        4520        4530        4540        4550        4560
TAGCAGGGAC AGTTATTCTT CTGATAGACA TTTATCACAA TACCATGATC ATCACAAAGA CAGGCATCAG GGAGATGCTT
      4570        4580        4590        4600        4610        4620        4630        4640
ACAAGAAAAG TGACTCCAGG AAAAGGCCAT ATTCAGCCTT CAGTAATGGA AAAGATCACA GAGACTGGGA TCACTACAAA
      4650        4660        4670        4680        4690        4700        4710        4720
CAGGACAGCA GATACTACAG TGATAGTAAA CATAGAAAGT TAGATGACCA CAGGAGCAGA GACCACAGGT CAAACCTGGA
      4730        4740        4750        4760        4770        4780        4790        4800
AGGAAACTTA AAAGACAGCC GGGGTCATTC AGATCACCGC TCCCATTCAG ACCACAGGAT ACACTCAGAT CACCGTTCCA
      4810        4820        4830        4840        4850        4860        4870        4880
CTTCAGAATA CAGCCATCAT AAATCTTCGA GAGATTATAG ATACCACTCA GACTGGCAAA TGGACCACAG AGCTTCTGGT
      4890        4900        4910        4920        4930        4940        4950        4960
AGTGGCCCGA GGTCACCACT AGATCAGAGG TCTCCTTATG GTTCAAGATC TCCCCTAGGA CACAGATCTC CATTTGAACA
      4970        4980        4990        5000        5010        5020        5030        5040
CTCATCAGAT CACAAAAGTA CACCTGAACA TACATGGAGT AGCCGGAAGA CATAACAAAG ACTGACATTT TCTGGACCTT
      5050        5060        5070        5080        5090        5100        5110        5120
CTTTTTAGCC ATATACAGTA AACTAACACA GTAATTGCCT TACATGACTT GAAAGATATG GACTGGATAT TCTATCAGTA
      5130        5140        5150        5160        5170        5180        5190        5200
GCAGTATTGT TACTTCTTTC CAGGATGCAA GGTCTATTAT CCCAACAGAA GAAAAATATT TTTGTATTTA AAGTTTATGC
      5210        5220        5230        5240        5250        5260        5270        5280
TGCACTGTGC TGCAAATGTT GTGGCACTTT TTTTTAAGA AATGGAAGAT GTTTACTTTT ACAGGGACCT CAACACTGCC
      5290        5300        5310        5320        5330        5340        5350        5360
CCTTTCAGAC TGGATCTTAC TATAAAACTC TTCATGTCAA AGTGGTTCTA GGCTGAACAC AGATTAAATT ATGTTTGTAA
      5370        5380        5390        5400        5410        5420        5430        5440
ATGAACACTT AAACACTGAC CTGTGCTTAT GTTTCAGGAA AGAATGGGGG ATTTATTTTG TTTTATTTCT TGGTAGAGAA
      5450        5460        5470        5480        5490        5500        5510        5520
CTCTCAAGGA CTTTGTTCAC TTTCCAAAGC TACTTGTTTA CATTGTACAC TGCGACCACC TTGCCGCTTT TCATCACAAG
      5530        5540        5550        5560        5570        5580        5590        5600
CTTGAATATT TAAATTCTGT ACCTACAGTT GTAAAATAGC CAGGATTTCT CCTGTTTGTG ATCAGTTATA ATGCCTTTTT
      5610        5620        5630        5640        5650        5660        5670        5680
ATGAAACAAA CAAACAAACA AAAAACAATT AAAAAAAAAA ACACAACAAA ACCAACAAAT GGCTGTAAAT TATTGTAAAT
      5690        5700        5710        5720        5730        5740        5750        5760
TAATTAAATG AGCTTTTTTC CGTCAGGCTT TTTTTGGCTG TTCCTTTCCC CAACAACTCA GGCCTTCTTT TCACAAAGTC
      5770        5780        5790        5800        5810        5820        5830        5840
```

```
AGTATACTTA CATGTTTTAA TAAAATATCT CGATGGAATC AGAATGTAAA AATGGGGAAG GGAATATTTT ATTCCATTTA
      5850       5860       5870       5880       5890       5900       5910       5920
>TGCTCCTTT TTTATTGGAT ACTTTTACAT ACCTGTTTTT GGTTGTTTTA TTTTATTTTT TTTTTCTATT AAACTGTCAG
      5930       5940       5950       5960       5970       5980       5990       6000
TGTTGTGATT GTTGTAATGA ACAGTGAGAA TATCCCACTC TAAACTGTGC CCTGGAAAGC TTTTCAGGTG CATTGGTTTA
      6010       6020       6030       6040       6050       6060       6070       6080
AAAGAAGGAA GTGTTCTATA GGTGAACACT TCAAAACCCA GATCAGCCAA GATTCATTGT AAATCCATTT GTTTTCCCTC
      6090       6100       6110       6120       6130       6140       6150       6160
TTTAACATGG GCAATAATGT CAAATGTGCT ATGCAGCAGT TAATATTTTA GAAGATTTGA ATGACTTTAT TAACAGAATT
      6170       6180       6190       6200       6210       6220       6230       6240
GTTACAATGC ACACTGATTG TACATAGATA ACTTCTATCT GACAAATTAA ATTAACTAAA ACCAAAAAAA ACC
      6250       6260       6270       6280       6290       6300       6310
```

Figure 6. The strategies used to determine the nucleotide sequence of *CHD-1A* and *CHD-W* given in fig. 5 and fig. 8. The top line represents the mouse clone given by [Delmas, 1993 #415]. The three 'Z' clones of *CHD-1A* and the 'CC4' and 'CC14' clones of *CHD-W* were derived from either a stage 10-12 or a 10 day chick cDNA library respectively. Arrows indicate the direction of sequence determination.

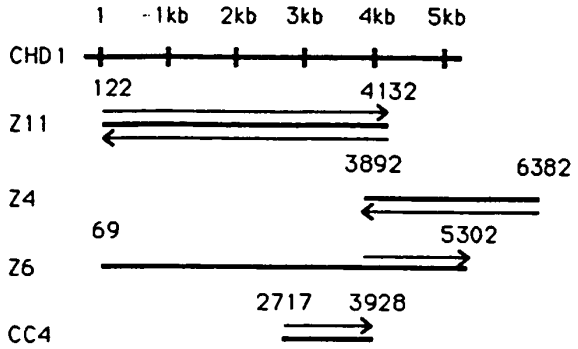


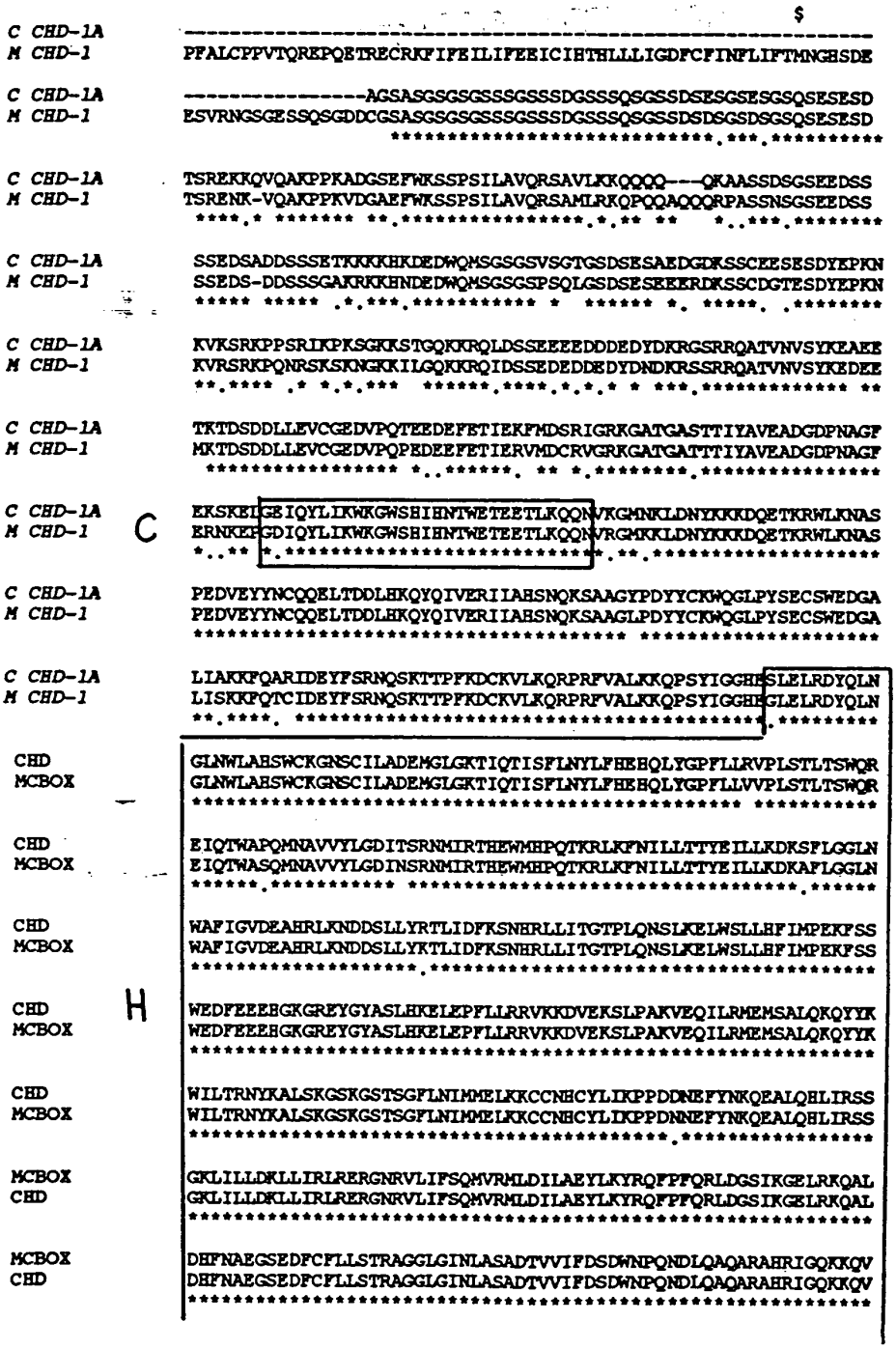

Figure 7. A composite nucleotide sequence and putative translation of the motif that is found spliced to a proportion of the 5' or 3' terminii of *CHD-1* clones or the 3' end of the *CHD-W* clone CC14.

```
                 D  E  I  V  S  V  K  H  L  H  K  K  I  K  T  E
CHD-1A     1  GATGAGATTGTTTCAGTGAAACATCTACATAAAAAAATAAAAACAGAAA
CHD-W      1  GATGGGATTGTTTCAGTGAAACATCCACATAAAAAAATAAAAGCAGAAA
                 D  G  I  V  S  V  K  H  P  H  K  K  I  K  A  E

                 K  E  N  E  E  K  P  E  P  D  I  G  I  K  K  E  A
CHD-1A    51  AAAGAAAATGAAGAAAAGCCTGAGCCAGATATTGGTATAAAGAAGGAAGCT
CHD-W     51  AAAGAAAATGAAGAAAAAGATGAGCCAGAGATTGGTATAAAGAAGGAAGCT
                 K  E  N  E  E  K  D  E  P  E  I  G  I  K  K  E  A

                 E  E  K  R  E  T  K  E  K  E  N  K  R  E  L  K  R
CHD-1A   101  GAAGAAAAAAGAGAGACAAAAGAGAAGGAAAATAAAAGGGAATTGAAAAGG
CHD-W    101  GGAGAAAAAAGAGAGACAAAAGAAAAGGAAAATAAGA
                 G  E  K  R  E  T  K  E  K  E  N  K

                 E  K  K  E  K  E  D  K  K  E  L  K  E  K  D  N  K
CHD-1A   151  GAGAAAAAAGAAAAAGAGGATAAGAAAGAATTAAAAGAAAAAGATAATAAA

                 E  K  R  E  N  K  V  K  E  S  T  Q  K  E  K  E  V
CHD-1A   201  GAAAAGAGAGAAAACAAAGTAAAAGAATCCACACAGAAAGAAAAAGAAGTG

                 K  E  E  K
CHD-1A   251  AAGGAAGAGAAG
```

Figure 8. A partial nucleotide sequence of *CHD-W* as defined by the clones CC4 and CC14.

```
ATTTATCGGC TAGTCACAAA AGGATCAGTA GAAGAAGATA TTCTTGAAAG AGCCAAGAAA AAGATGGTGT TAGATCATTT
        10         20         30         40         50         60         70         80
AGTGATTCAG AGAATGGACA CCACAGGGAA AACTGTACTA CATACAGGCT CTACTCCTTC AAGCTCAACA CCTTTTAATA
        90        100        110        120        130        140        150        160
AGGAAGAGTT ATCAGCAATT TTGAAGTTTG GTGCTGAGGA ACTTTTTAAA GAACCTGAAN NNGAAGAAGA GGAGCCTCAG
       170        180        190        200        210        220        230        240
GAGATGGATA TAGATGAAAT CCTGAAGAGG NCTGAAACTC GAGAAAATGA GTCAGGCCCA TTAACTGTAG GAGATGAGTT
       250        260        270        280        290        300        310        320
ACTTTCACAG TTCAAGGTAG CTAACTTTTC CAATATGGAT GAAGATGACA TTGAATTGGA ACCAGAACAA AATCTAAGAA
       330        340        350        360        370        380        390        400
ACTGGGAAGA AATCATTCCA GAAGTTCAGT GGCGACGAAT AGAGGGGNNG GAAAGACAAA AAGAACTTGA AGAAATATAT
       410        420        430        440        450        460        470        480
ATGCTTCCAA GAATGAGAAA CTGTGCAAAA CAGATCAGCT TTAATGGAAA TGAAGGGAGA TGCAGTAGGA GCAGAAGATA
       490        500        510        520        530        540        550        560
TTCTGGATCT GATAGTGATT CCATCTCAGA AAGAAAACGA CCAAAAAAAC GTGGACGACC ACGAACTATT CCCCGTGAAA
       570        580        590        600        610        620        630        640
ACATTAAAGG ATTTAGTGAT GCAGAGATTA GACGATTTAT CAAGAGTTAC AAGAAATTTG GTGGCCCAGT TGAAAGGTTA
       650        660        670        680        690        700        710        720
GATGCTATAG CTAGAGATGC TGAGCTAGTT GATAAATCTG AAACAGACCT TAGACGTCTG GGAGAACTTG TACATAATGG
       730        740        750        760        770        780        790        800
ATGCATTAAG GCTTTAAATG ATAATGACTT TGGTCAAGGA AGAACAGGTG GTAGATTTGG GAAAGTTAAA GGCCCAACAT
       810        820        830        840        850        860        870        880
TCCGAATAGC AGGAGTGCAG GTGAATGCAA AGCTAGTCAT TTCTCACGAA GAAGAGTTGG CACCATTGCA TAAATCGATT
       890        900        910        920        930        940        950        960
CCTTCAGATC CAGAAGAAAG GAAAAGATAT GTCATCCCAT ACCACACCAA AGCAGCTCAT TTTGATATAG ATTGGGGTAA
       970        980        990       1000       1010       1020       1030       1040
AGAAGATGAT TCCAATCTGT TAATAGGCAT CTATGAATAT GGTTATGGCA GTTGGGAAAT GATAAAAATG GATCCTGATC
      1050       1060       1070       1080       1090       1100       1110       1120
TCAGTTTGAC ACAGAAGATT TTACCTGATG ATCCAGATAA GAAACCCCAG GCTAAGCAGT TACAGACTCG TGCAGATTAC
      1130       1140       1150       1160       1170       1180       1190       1200
CTCATTAAAT TACTGAATAA AGACCTTGCA AGAAAGGAAG CACAGAGACT TGCTGGTGCA GGCAATTCAA AGAGGAGAAA
      1210       1220       1230       1240       1250       1260       1270       1280
AACAAGAAGT AAGAAGAATA AAGCAACAAA GGCTGCAAAA AAAAAAAAAA AAAAA
      1290       1300       1310       1320       1330
```

Figure 9. An alignment of the deduced amino acid sequences of the chicken (C) *CHD-1A* and mouse (M) *CHD-1*. With gaps introduced to maximize alignment they show a sequence identity of 91.8% over 1339 residues. The $ sign indicates start and stop codons. Boxed sections are the chromodomain (C), Helicase, (H) and the region containing the DNA binding domain (B) identified by Delmas *et al.*, (1993). A trimer repeat of a basic HSDHR motif is underlined. A* denotes residue identity and . similarity.



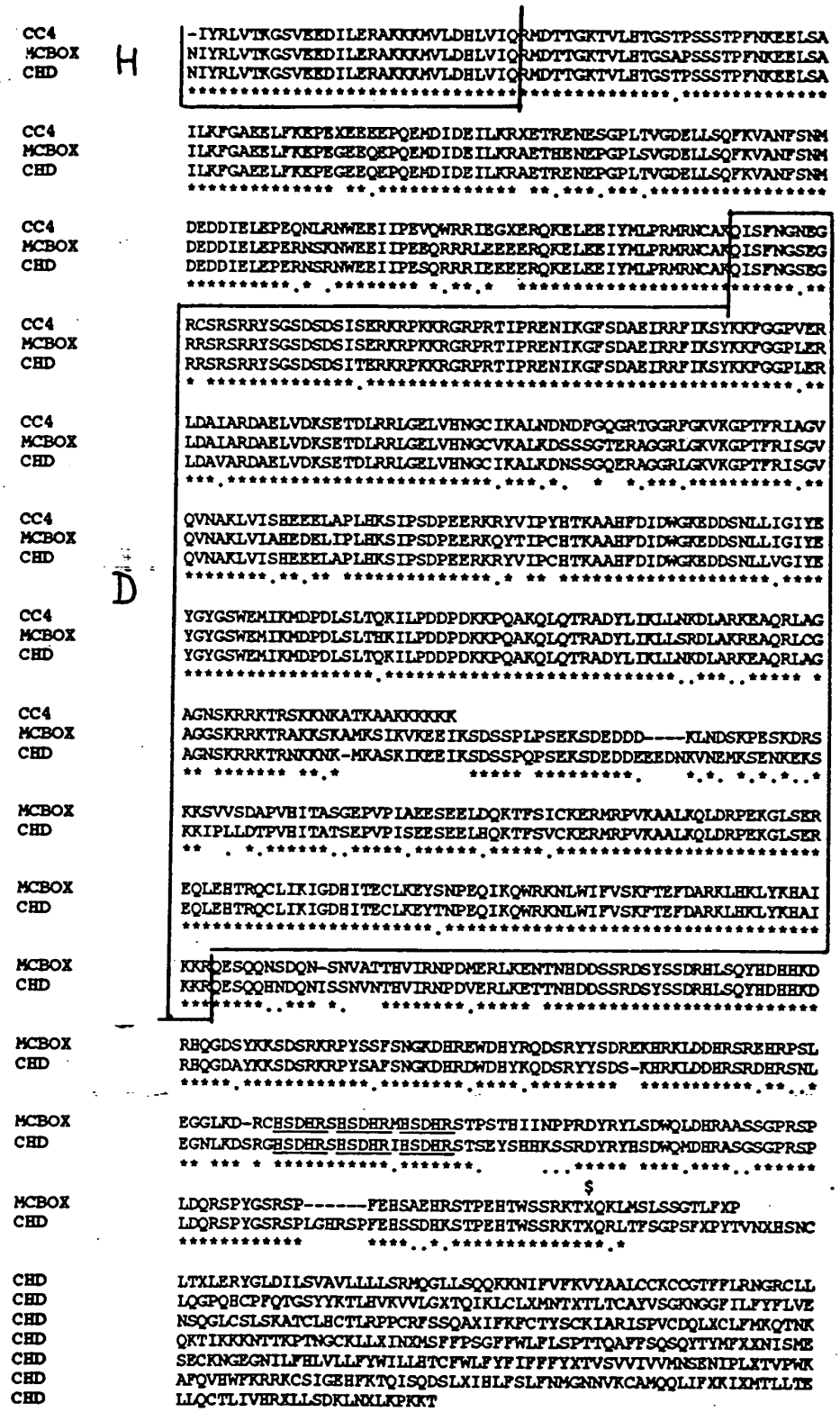

CC4
MCBOX
CHD
H
D

Figure 10. An alignment of the deduced amino acid sequences of *CHD-1A* and *CHD-1Y* a putative yeast homologue of the chicken gene identified through a search of the EMBL data base. With gaps introduced to maximize alignment they show a sequence identity of 37.7% over 1538 residues. | indicates identity and : conservative substitution.

```
  55 KPPKADGSEFWKSSPSILAVQRSAVLKKQQQQQKAASSDSGSEEDSSSSE 104
     .::|. :.|.:.. |.:.::.||   .:...:|..  .||:.|:|    |
2654 MAAKDISTEVLQN.PELYGLRRS...HRAAAHQQNYFNDSDDEDD....E 2695

 105 DSADDSSSETKKKKHKDEDWQMSGSGSVSGTGSDSESAEDGDKSSCEESE 154
     |. .:|... ... ..|||        ...:.::::.:|:||:|... ||.:
2696 DNIKQSRRKRMTTIEDDED.......EFEDEEGEEDSGEDEDEEDFEEDD 2738

 155 SDYEPKNKVKSRKPPSRIKPKSGKKSTGQKKRQLDSSEEEEDDDEDYDKR 204
     . |:.. | ...||.||.|.|| .|..:|..:| . . .        .:
2739 DYYGSPIKQNRSKPKSRTKSKSKSKPKSQSEKQSTVKIP........TRF 2780

 205 GSRRQATVNVSYKEAEETKTDSDDLLE....VCGEDVPQT....EEDEFE 246
     :.|.. ||| ... .::.  :|:| .:    :::|:|.:.    :.::|.
2781 SNRQNKTVNYNIDYSDDDLLESEDDYGSEEALSEENVHEASANPQPEDFH 2830

 247 TIEKFMDSRIGRKGATGASTTIYAVEADGDPNAGFEKSKELGEIQYLIKW 296
     .|: .:: |:         .|.: .... :.. ::::.:||  :.::||||
2831 GIDIVINHRL........KTSLEEGKVLEKTVPDLNNCKE..NYEFLIKW 2870

 297 KGWSHIHNTWETEETLKQQNVKGMNKLDNYKKK...DQETKRWLKNASPE 343
     .: ||:|||||| |.: |  |:|:.:|||| |.   :::  |: . ..:|
2871 TDESHLHNTWETYESIGQ..VRGLKRLDNYCKQFIIEDQQVRLDPYVTAE 2918

 344 DVEYYNCQQELTDDLHKQYQIVERIIA..HSNQKSAAGYPDYYCKWQGLP 391
     |:|. : :.|   |  .::::.|||.  :.. ..:.:  :|..||. |
2919 DIEIMDMERERRLDEFEEFHVPERIIDSQRASLEDGTSQLQYLVKWRRLN 2968

 392 YSECSWEDGALIAKKFQARIDEYFSRNQSKTTPFKDCKVLKQRPRFVALK 441
     |.|..||::. |.|  ...:..: .|:.||. |  .:.  .|||||  |.
2969 YDEATWENATDIVKLAPEQVKHFQNRENSKILPQYSSNYTSQRPRFEKLS 3018

 442 KQPSYIGGHESLELRDYQLNGLNWLAHSWCKGNSCILADEMGLGKTIQTI 491
      ||.:| |    ||||:||.|:||:|  |:||:..|||||||||||:||:
3019 VQPPFIKGG...ELRDFQLTGINWMAFLWSKGDNGILADEMGLGKTVQTV 3065

 492 SFLNYLFHEHQLYGPFLLRVPLSTLTSWQREIQTWAPQMNAVVYLGDITS 541
     .|:.:|: .:.  || :: |||||:..|  .::.|||::|.:.|:|: .|
3066 AFISWLIFARRQNGPHIIVVPLSTMPAWLDTFEKWAPDLNCICYMGNQKS 3115

 542 RNMIRTHEW....MHPQTKVLKFNILLTTYEILLKDKSFLGGLNWAFIGV 587
     |: ||..|:    . ...| :|||:||||||.:|||:. ||::.|.|::|
3116 RDTIREYEFYTNPRAKGKKTMKFNVLLTTYEYILKDRAELGSIKWQFMAV 3165

 588 DEAHRLKNDDSLLYRTLIDFKSNHRLLITGTPLQNSLKELWSLLHFIMPE 637
     ||||||||.:| || .| .|| .:|:|||||||||.:||| .|::|:||:
3166 DEAHRLKNAESSLYESLNSFKVANRMLITGTPLQNNIKELAALVNFLMPG 3215

 638 KFSSWEDFE.EEHGKGREYGYASLHKELEPFLLRRVKKDVEKSLPAKVEQ 686
     :|.  :::: |:::.:.|  . .||: ::||:|||:||||||||.|.|.
3216 RFTIDQEIDFENQDEEQEEYIHDLHRRIQPFILRRLKKDVEKSLPSKTER 3265

 687 ILRMEMSALQKQYYKWILTRNYKALSKGSKGSTSGFLNIMMELKKCCNHC 736
     |||:|:|.:|.:||| |||:||.||. |.||:  ::|||| ||||.:||.
3266 ILRVELSDVQTEYYKNILTKNYSALTAGAKGGHFSLLNIMNELKKASNHP 3315

 737 YLIKPPDDNEF.......YNKQEALQHLIRSSGKLILLDKLLIRLRERGN 779
     ||:. :::. :        .:::.|. ||.||||::|||.||.||:. |:
3316 YLFDNAEERVLQKFGDGKMTRENVLRGLIMSSGKMVLLDQLLTRLKKDGH 3365
```

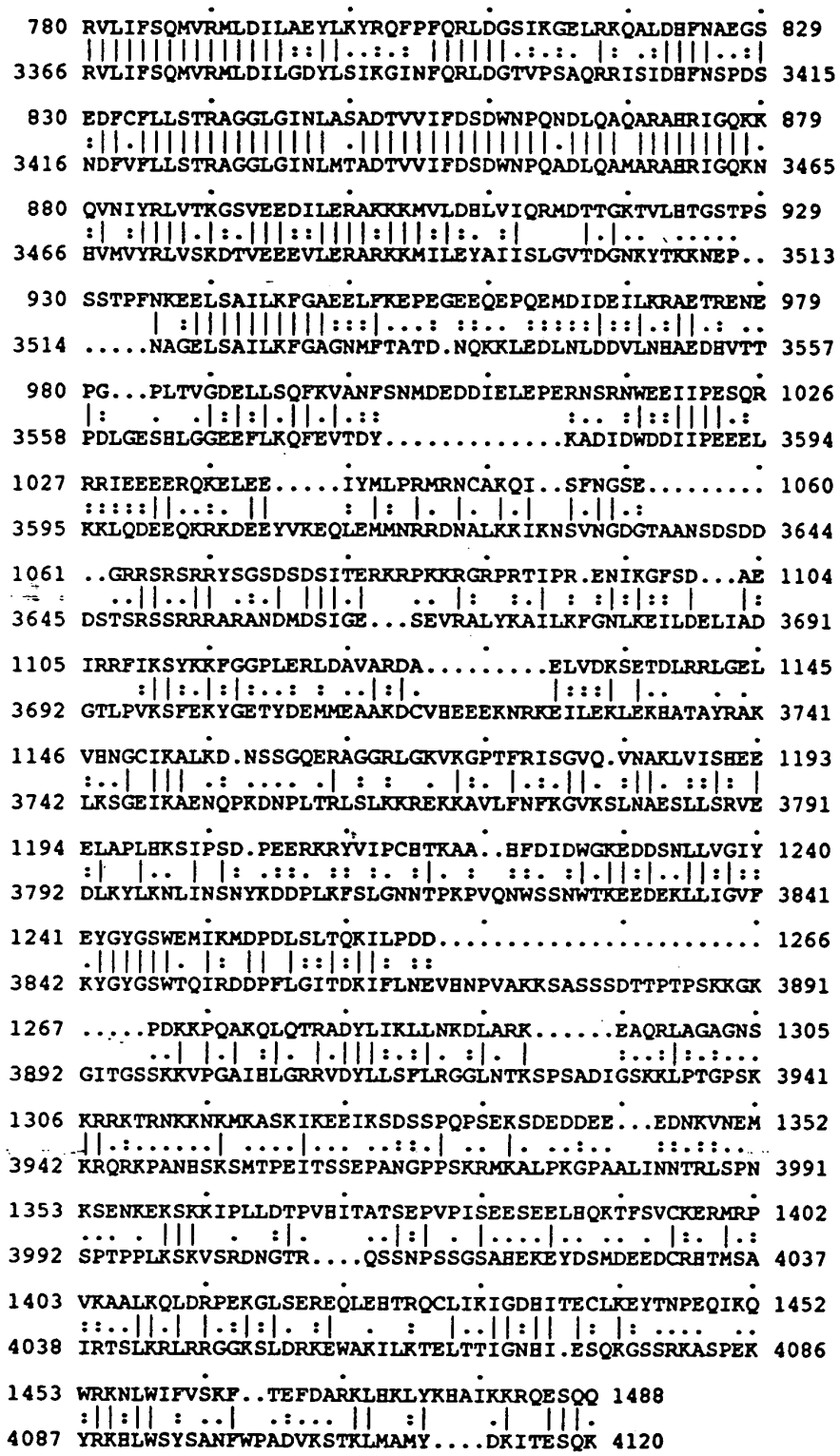
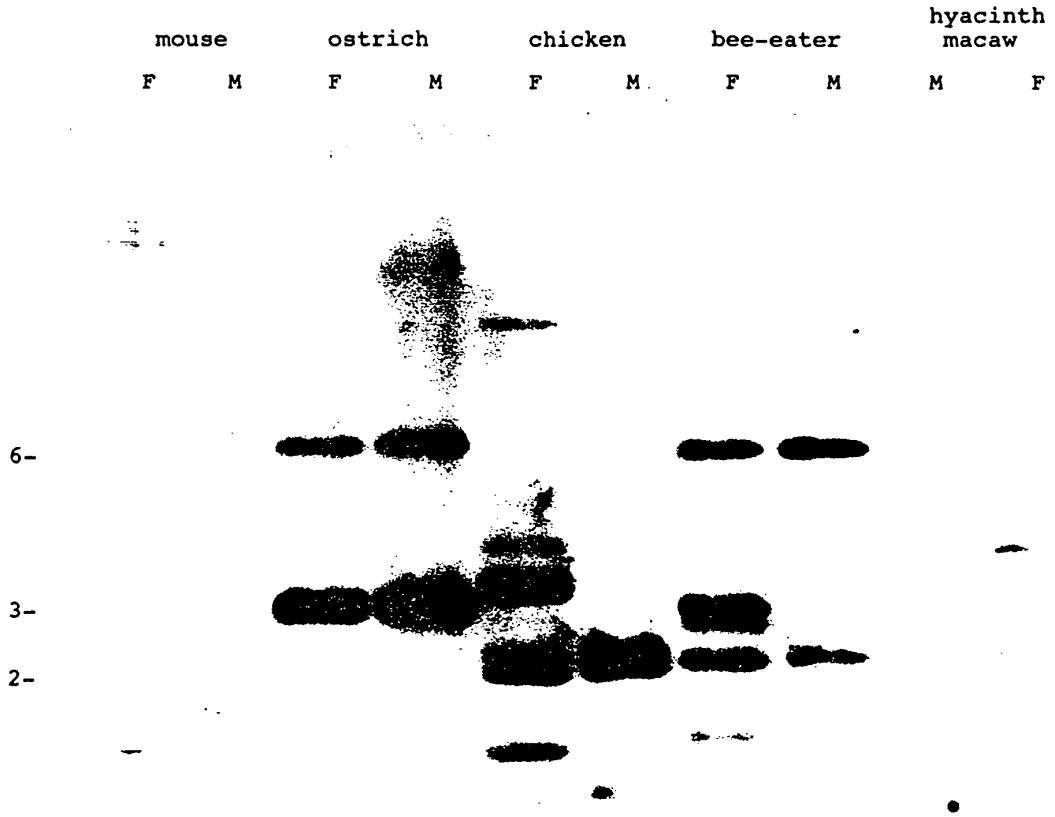


Figure 11. Comparison of 9 chromodomain sequences. Vertical lines indicate the extent of the chromodomain as defined by [Paro, 1991 #457]. The top three sequences represent the CHD class of chromodomain to add to the HP1 class and Pc classes as defined by [Pearce, 1992 #424]. The first letter of each annotation indicates the animal of origin: C, chicken; M mouse; D, Drosphila; H, human; Y, *S. cerivisiae* whilst the remainder identifies the gene type. The yeast gene is a possible CHD homologue identified by its close identity to the vertebrate forms. * indicates sequence identity within the groups and ^ identity between all nine sequences. _ indicate amino acid residues inside and downstream of the motif that are characteristic of the CHD class chromobox.

```
CCHD   AVEAD| GDPNAGFEKSKELGE.IQYLIKWKGWSHIHNTWETEET| LKQQNVKGMNKLDNYKK
MCHD   AVEAD| GDPNAGFERNKEPGD.IQYLIKWKGWSHIHNTWETEET| LKQQNVRGNKKLDNYKK
YCHD   EGKVL| EKTVPDLNNCKE..N.YEFLIKWTDESHLHNTWETYES| IGQ..VRGLKRLDNYCK
                    __          ____    __ ______ *     *  _  _  _    ____ _
DHP1   EEEEE| YAVEKIIDRRVRKGK.VEYYLKWKGYPETENTWEPENN| LDCQDLIQQY
HHP1   EDEEE| YVVEKVLDRRVVKGKQVEYLLKWKGFSEEHNTWEPEKN| LDCPELISEF
MMOD1  EEEEE| YVVEKVLDRRVVKGK.VEYLLKWKGFSDEDNTWEPEEN| LDCPDLIAEF
MMOD2  AEPEE| FVVEKVLDRRVVNGK.VEYFLKWKGFTDADNTWEPEEN| LDCPELIEDF
         **|   ***  ****  ** *** *****     ****** *| ***  **
DPC    PVDLV| YAAEKIIQKRVKKGV.VEYRVKWKGWNQRYNTWEPENN| ILDRRLIDIY
MMOD3  VGEQV| FAAECILSKRLRKGK.LEYLVKWRGWSSKHNSWEPEEN| ILDPRLLLAF
          *  *** *  **  ***  ** *** **    * **** * ******
                                ^^       ^ ^^
```

Figure 12. Genomic Southern blots of DNA from male and female chickens and lesser black-backed gulls digested with PvuII and probed with a 433bp HindIII/Sac fragment of pGT8 (fig 4.) at moderate stringency. Hybridization with female linked fragments and fragments common to both sexes can be observed in both species. Numbers give approximate sizes in kilobases.

Figure 13. Genomic Southern blots of DNA from male (M) and female (F) mice, ostrich, chicken, bee-eater and hyacinth macaw probed with the 1335bp insert of CC4 at moderate stringency. Hybridization with mouse and ostrich is with fragments shared by both sexes whilst the non-ratite birds show additional hybridization to female specific fragments. In these latter species, the signal from female linked hybrids is stronger than with autosomal/Z linked fragments indicating that the probe is derived from the W chromosome. Numbers give approximate sizes in kilobases.



PCT/GB 96/01341

5 6 96

Stevens Hewlett & Perkins